PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 :
C12N 15/86, C07K 14/19, C12N 7/01, 7/04, A61K 48/00, 39/20

A1

(11) International Publication Number: **WO 99/61637**

(43) International Publication Date: 2 December 1999 (02.12.99)

(21) International Application Number: PCT/CA99/00479

(22) International Filing Date: 21 May 1999 (21.05.99)

(30) Priority Data:
2,238,660 22 May 1998 (22.05.98) CA

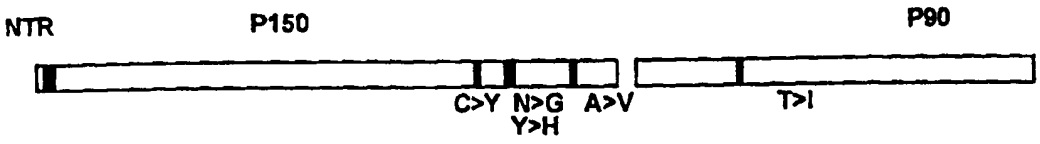
(71) Applicant *(for all designated States except US)*: THE UNIVERSITY OF BRITISH COLUMBIA [CA/CA]; IRC Room 331, 2194 Health Sciences Mall, Vancouver, British Columbia V6T 1Z3 (CA).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: CHANTLER, Janet [CA/CA]; 6158 Alma Street, Vancouver, British Columbia V6N 1Y6 (CA). LUND, Karen [CA/CA]; Apartment 404, 1122 Gilford Street, Vancouver, British Columbia V6G 2P5 (CA).

(81) Designated States: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: GENE SEQUENCES OF RUBELLA VIRUS ASSOCIATED WITH ATTENUATION



I Cendehill mutations designated by single letter code

(57) Abstract

The complete sequence of the Cendehill rubella strain is provided, including infectious cDNA clones derived therefrom. Portions of the Cendehill genome responsible for decreased arthritogenicity and immunogenicity are identified. Modified rubella cDNA, RNA and virus is provided incorporating Cendehill and non–Cendehill sequences.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	**ES**	Spain	**LS**	Lesotho	**SI**	Slovenia
AM	Armenia	**FI**	Finland	**LT**	Lithuania	**SK**	Slovakia
AT	Austria	**FR**	France	**LU**	Luxembourg	**SN**	Senegal
AU	Australia	**GA**	Gabon	**LV**	Latvia	**SZ**	Swaziland
AZ	Azerbaijan	**GB**	United Kingdom	**MC**	Monaco	**TD**	Chad
BA	Bosnia and Herzegovina	**GE**	Georgia	**MD**	Republic of Moldova	**TG**	Togo
BB	Barbados	**GH**	Ghana	**MG**	Madagascar	**TJ**	Tajikistan
BE	Belgium	**GN**	Guinea	**MK**	The former Yugoslav Republic of Macedonia	**TM**	Turkmenistan
BF	Burkina Faso	**GR**	Greece			**TR**	Turkey
BG	Bulgaria	**HU**	Hungary	**ML**	Mali	**TT**	Trinidad and Tobago
BJ	Benin	**IE**	Ireland	**MN**	Mongolia	**UA**	Ukraine
BR	Brazil	**IL**	Israel	**MR**	Mauritania	**UG**	Uganda
BY	Belarus	**IS**	Iceland	**MW**	Malawi	**US**	United States of America
CA	Canada	**IT**	Italy	**MX**	Mexico	**UZ**	Uzbekistan
CF	Central African Republic	**JP**	Japan	**NE**	Niger	**VN**	Viet Nam
CG	Congo	**KE**	Kenya	**NL**	Netherlands	**YU**	Yugoslavia
CH	Switzerland	**KG**	Kyrgyzstan	**NO**	Norway	**ZW**	Zimbabwe
CI	Côte d'Ivoire	**KP**	Democratic People's Republic of Korea	**NZ**	New Zealand		
CM	Cameroon			**PL**	Poland		
CN	China	**KR**	Republic of Korea	**PT**	Portugal		
CU	Cuba	**KZ**	Kazakstan	**RO**	Romania		
CZ	Czech Republic	**LC**	Saint Lucia	**RU**	Russian Federation		
DE	Germany	**LI**	Liechtenstein	**SD**	Sudan		
DK	Denmark	**LK**	Sri Lanka	**SE**	Sweden		
EE	Estonia	**LR**	Liberia	**SG**	Singapore		

GENE SEQUENCES OF RUBELLA VIRUS ASSOCIATED WITH ATTENUATION

Background of the Invention

Rubella virus is the causative agent of German measles, a viral infection associated with a mild fever and rash. The most serious complications of rubella occur during pregnancy due to transplacental passage of the virus to the fetus resulting in the widespread manifestations of congenital rubella. These include fetal loss, or multisystem defects in the newborn such as cataracts, deafness, cardiac abnormalities and microcephaly.

To prevent congenital infection, a universal vaccination scheme for all children around 15 months of age was implemented in North America in 1969, using attenuated vaccines which had recently been developed. While reducing the level of rubella circulating in the community, vaccination of young children did not significantly alter the proportion of women entering their childbearing years without protective levels of circulating antibody - reported to be around 10-15%. This population was therefore also targeted for vaccination.

Vaccination reduced the incidence of congenital rubella but was found to be associated with a number of sequelae, particularly in women over 25 years of age. Symptoms included arthritis, neurological manifestations and chronic fatigue. The most notable complication of rubella immunisation was arthritis which has also frequently been documented as a consequence of natural rubella. The joint symptoms induced can be severe in the acute stage but usually resolve without causing permanent joint damage. Occasionally, however, chronic or recurrent arthritis develops which can persist for many months or years in certain individuals (Ford et al., 1988)

Several vaccines have been used in North America since 1969. These include two variants of the HPV77 strain

originally produced by Dr. H. Meyer from the wild strain M33 by multiple passages in monkey kidney cells (Meyer et al., 1969). The HPV77 strain was further attenuated by a further 5 passages in duck embryo cells (to give the HPV77/DE5 strain) or by 12 passages in dog kidney cells (to give the HPV77/DK12 strain). The HPV77/DK12 vaccine proved to be too reactogenic even in children and was soon removed from distribution. The HPV77/DE5 vaccine was used as part of the M-M-RI vaccine (measles/mumps/rubella combined vaccine; Merck Sharp & Dohme; West Point, Pa. U.S.A.) until 1979 when it was replaced in the M-M-RII vaccine by the RA27/3 strain (Plotkin and Buser, 1985), which is the current vaccine strain used in North America.

The Cendehill strain (Peetermans & Huygelen, 1967) was developed in Belgium and was the predominant strain used in vaccine production in Europe until 1989. The Cendehill strain is reported to be associated with a decreased incidence of complications in the adult female population in a comparative study of five vaccines. Best et al. (1974) reported that acute arthritis occurred in only 3% of individuals immunised with Cendehill vaccine but in 17% of those receiving RA27/3. Moreover the symptoms with RA27/3 were also more prolonged. The disadvantage of the Cendehill vaccine was that the mean titre of HAI antibody induced in vaccine recipients was lower than that obtained with the RA27/3 strain indicating that Cendehill is less immunogenic.

A close correlation has been found between the ability of a given strain of rubella virus to infect and persist in human joint tissue in culture and its association with the induction of arthropathy in vivo, suggesting that tropism for joint tissue is an important determinant of the ability to induce joint symptoms (arthritogenicity). As reported in Miki and Chantler (1992), wild-type strains (Therien and M33) were found to grow to high titres of 10^6-10^7 pfu/ml in

the medium of either cell cultures or organ cultures derived from human joint tissue. In contrast the RA27/3 strain was considerably restricted for growth giving yields of 10^3-10^4 pfu/ml and the Cendehill strain showed no growth at all. These results correlate with the known associations of rubella strains and joint symptoms in vivo.

Rubella virus is a small (60-70 nm) enveloped togavirus, the sole member of the genus *Rubivirus*. It has a single-stranded RNA genome approximately 10kb in size. The genomic RNA is positive-stranded which means that it can act as mRNA within the infected cell. The sequence of the entire genome has been determined for two wild-type strains Therien and M33 (Dominguez et al., 1990; Gillam et al., 1993, Genbank No. X72393), and the RA27/3 vaccine strain (Pugachev et al., 1997). The genome contains two large open-reading frames (ORF's) which code for the structural proteins (3' proximal 3189 nucleotides) and non-structural proteins (5' proximal 6345 nucleotides). The current understanding is that the open-reading frames for the structural and the non-structural proteins are separated by a region of about 123 nucleotides.

The infected cell contains two virus-induced positive-strand RNA species, the genomic RNA (40s; 10kb) and a sub-genomic mRNA (26s; 3kb) which encodes the major ORF for the structural proteins. The ORF for structural proteins is translated into a 110kd polyprotein and is subsequently cleaved by cellular signal peptidase into the three structural viral proteins, E1, E2, and C. The order of structural genes was originally determined by synchronised translation as being NH_2-C-E2-E1-COOH, which was confirmed by sequence analysis of cDNA clones of the subgenomic mRNA (Clarke et al., 1987; Frey & Marr, 1988; and Zheng et al., 1989).

The non-structural (NS) genes are translated from the full-length genomic RNA as a >200kD polyprotein which is subsequently cleaved into two non-structural proteins, p150 and p90. These comprise the enzymes required for viral replication in the cell. Protein p150, nearest the 5' terminus, is 1300 amino-acids in length and encodes the putative methyltransferase function and the viral protease. Protein p90 is 905 amino-acids long and has regions of homology with global helicase and replicase domains.

Summary of the Invention

This invention provides nucleic acids (DNA or RNA) comprising one or more sequences of nucleotides corresponding to all or part of the genome of the Cendehill strain of rubella virus. Nucleic acids of this invention may encode an infectious virus of the Cendehill strain or one having an attenuated phenotype equivalent to Cendehill strain. DNA of this invention may be in a plasmid or viral vector which enables replication and/or transcription of the Cendehill cDNA and is referred to herein as a Cendehill infectious clone. The infectious clone may be used to produce a DNA vaccine for rubella virus.

This invention also provides a nucleic acid (DNA or RNA) comprising a sequence of nucleotides that includes a first portion corresponding to one or more of the non-translated regions, p150, p90, C, E2 and E1 gene regions of Cendehill strain and a second portion that is derived from another rubella virus strain such that the product encodes a novel infectious chimeric rubella virus strain. DNA of this invention may be in a plasmid or viral vector forming an infectious clone.

This invention also provides a chimeric Cendehill/RA27/3 clone whose genome includes a first portion corresponding to the Cendehill 5' non-translated

RNA, Cendehill p150 and p90 and wherein a second portion corresponds to the structural gene region and the 3' non-translated region of RA27/3 strain. This clone can be used to produce a chimeric virus that expresses the structural proteins of RA27/3 but has the genetic structure at the 5'end and in the non-structural genes of Cendehill strain that determine the non-arthrotropic nature of this strain.

This invention also provides RNA encoding the entire genome of Cendehill or the Cendehill/RA27/3 chimera or a fragment thereof, by transcribing the aforementioned DNA. This invention also provides rubella virus produced by transcribing the DNA, transfecting cells with the RNA so derived, and recovering virus from cells so transfected.

This invention also provides a nucleic acid encoding one or more Cendehill strain rubella virus proteins selected from the group consisting of: p150, p90, C, E1 and E2, or wherein the nucleic acid corresponds to a non-translated region of the Cendehill genome. The nucleic acid may be DNA or RNA and may be incorporated into a plasmid or viral vector for expression of protein.

This invention also provides a method of producing Cendehill viral protein comprising the steps of expressing a DNA sequence encoding a protein corresponding to Cendehill protein p150, p90, C, E2 or E1 in a cell by means of a suitable expression vector and recovering the protein so expressed. The protein may be a Cendehill protein having a sequence corresponding to a portion of the cDNA sequence in Appendix 1 or the protein may be altered by modification of the Cendehill cDNA, as described herein.

This invention also provides a method of producing a recombinant DNA encoding a mutated or chimeric rubella virus exhibiting the lack of arthrotropicity of the

Cendehill strain but with additional advantageous properties that include, but are not restricted to, increased immunogenicity or stability of another rubella strain.

This method comprises steps whereby;

(a) nucleotides in Cendehill cDNA encoding viral structural proteins are altered such that the protein so encoded increases the immunogenicity or stability of a recombinant rubella virus comprising said protein; or

(b) nucleotides in the non-translated regions or non-structural gene region of cDNA for rubella virus other than Cendehill are altered to decrease arthritogenicity of a recombinant rubella virus coded for by the altered cDNA.

cDNA from steps (a) or (b), may be incorporated into a plasmid or viral vector to produce an infectious clone, from which RNA may be transcribed and transfected into cells to provide virus that may be used as a recombinant rubella vaccine. Alternatively, cDNA from (a) or (b) in a suitable vector may be used as a DNA vaccine.

This invention also provides a rubella virus whose genetic material comprises a first portion corresponding to one or more RNA sequences selected from the group consisting of: Cendehill non-translated RNA, Cendehill p150, p90, C, E1 and E2 RNA; and wherein a second portion of the genome corresponds to RNA of a rubella virus other than Cendehill.

This invention also provides a Cendehill viral protein free of virus, selected from the group consisting of: p150, p90, C, E1 and E2, produced by expressing Cendehill cDNA encoding said protein from an expression vector.

This invention also provides rubella cDNA, RNA or a rubella virus having one or more of the Cendehill strain-specific nucleotides selected from a group consisting of: 37-C, 55-G, 118-T(or U), 358-C, 2829-A, 3060-G, 3164-C,3528-T (or U), 4530-T (or U), 6611-C, 6770-G, 6771-G, 7428-T (or U), 8786-G, 8788-T (or U), 8864-A, 9180-T (or U), 9254-A, and 9741-T (or U). The aforesaid nucleotide numbers are in reference to nucleotides bearing the same numbers as shown in Appendix 1 for Cendehill. cDNA, RNA or virus of this invention may have the strain-specific nucleotide at a different nucleotide position number as compared to Cendehill, providing the context of the strain-specific nucleotide is the same as for Cendehill. In this instance, context defines the five nucleotides on either side of the strain-specific nucleotide in Cendehill.

This invention also provides a Cendehill cDNA, and genomic RNA that encodes a rubella virus protein selected from the group of proteins p150, p90, C, E1 and E2 and with one or more Cendehill strain-specific amino-acids defined as p150/929/tyr, p150/1006/gly, p150/1041/his, p150/1162/val, p90/1496/ile, C/ 34/pro, C/87/gly, E2/306/val, E2/413/ile, E1/759/asp, E1/785/met/, E1/890/leu, and E1/915/thr. The aforesaid strain-specific amino acids are identified by protein name, amino-acid position within the Cendehill rubella polyprotein, and the identity of an amino-acid at such a position. Such proteins of this invention include proteins having the strain-specific amino acid at a different amino acid position number in the protein as compared to Cendehill providing the context of the strain-specific amino acid is the same as for Cendehill. In this instance, context is defined as including the three amino acids to either side of the strain-specific amino acid in Cendehill. In this specification, reference to a strain-specific amino acid such as p150/929/tyr will be used to identify the amino

acid as well as a protein (eg. p150) containing the strain-specific amino acid, in context as described herein.

This invention also provides a nucleic acid (eg. DNA) for the first 5' non-translated region (NTR) and first stem loop (nucleotides 1 to 65) equivalent to that found in the Cendehill strain and characterised as being a major determinant of growth restriction in joint tissue. Specific characteristics of this stem loop in Cendehill include two nucleotide changes from the wild-type Therien strain, a U to C at nucleotide 37 in the predicted terminal loop that alters the size of the loop from 6 to 11 nucleotides, and an A to G at nucleotide 55 that increases the size of the predicted medial loop from 6-10 nucleotides. These two nucleotide changes at these positions and in the context found in Cendehill strain (defining the five nucleotides on either side of each nucleotide) are determinants of arthrotropism. Other mutations between nucleotides 20-28 and 52-60 that either increase or decrease the predicted size of the medial loop are included within the scope of this invention. Similarly any mutation that alters the predicted size of the terminal loop and alters the phenotypic characteristics of the virus are within the scope of this invention. Factors that define the determinants of joint cell restriction include sequence-specific changes in the medial or terminal loop or changes that alter the size of either or both of the loops. These regions include nucleotides 20-28, 33-43 and 52-60.

Appendix 1 sets out the sequence of cDNA representing the Cendehill genome. Location of the various non-translated regions and coding regions are shown. Two polyproteins are encoded, beginning at the start codons indicated for p150 and the C protein, respectively. The

amino acid sequence of each polyprotein and the respective structural and non-structural proteins may be determined from the nucleotide sequence of Appendix 1. In this specification, the location of an amino-acid will be given by reference to a residue number of a polyprotein, which residue number may be determined directly from the series of codons shown in Appendix 1 commencing at one or the other of the start codons.

The term "corresponding" as used in this specification means that when a nucleic acid, peptide or protein is described by reference to a specified nucleic acid, peptide or protein, the nucleic acid, peptide or protein so described may include a nucleotide or amino acid sequence which differs from the sequence of the specified nucleic acid, peptide or protein. Corresponding nucleic acids, polypeptides or proteins will include sequences of differing length or which differ by one or more substitutions, additions or deletions. Nucleic acids, peptides and proteins of this invention include fragments of specified nucleic acids, peptides or proteins and may include additional amino acid or nucleotide sequences from that specified. Furthermore, corresponding nucleic acids include complementary nucleic acids, meaning those nucleic acids capable of base pairing with a specified nucleic acid. Nucleic acids having sequences which differ from the sequence of a specified nucleic acid due to degeneracy of the genetic code are also included within the meaning of the term "corresponding". Further, nucleic acids which encode peptides or proteins in which there are conservative substitutions, additions or deletions as compared to a specified peptide or protein are included. Any and all such nucleotide variations and resulting amino acid polymorphisms which provide the advantages of this invention as described herein are within the scope of this invention.

Nucleic acids within the scope of this invention may contain linkers, modified or unmodified restriction endonuclease sites and other sequences of nucleotides useful for cloning, expression, or purification. Nucleic acids within the scope of this invention may be incorporated in a larger sequence of nucleotides, including plasmids and vectors useful for manipulation or expression of nucleic acids.

One measure of "correspondence" of nucleic acids, peptides or proteins with respect to this invention is relative "identity" between sequences. In the case of peptides or proteins, or in the case of nucleic acids defined according to a encoded peptide or protein correspondence includes a peptide having at least about 50% identity, more preferably at least about 70% identity, even more preferably at least about 90% identity, even more preferably at least about 95% and most preferably at least about 98-99% identity to a specified peptide or protein. Preferred measures of identity as between nucleic acids is the same as specified above for peptides with at least about 90% or at least about 98-99% identity being more or most preferable.

The term "identity" as used herein refers to the measure of identity of sequence between two peptides or between two nucleic acid molecules. Identity can be determined by comparing a position in each sequence which may be a line for purposes of comparison. Two amino acid or nucleic acid sequences are considered substantially identical if they share at least about 75% sequence identity, preferably at least about 90% sequence identity, even more preferably at least 95% sequence identity and most preferably at least about 98-99% identity.

Sequence identity may be determined by the BLAST algorithm described in Altschul et al. (1990) J. Mol. Biol.

215:403-410, using the published default settings. When a position in the compared sequence is occupied by the same base or amino acid, the molecules are considered to have shared identity at that position. The degree of identity between sequences is a function of the number of matching positions shared by the sequences.

An alternate measure of identity of nucleic acid sequences is to determine whether two sequences hybridize to each other under low stringency, and preferably high stringency conditions. Such sequences are substantially identical when they will hybridize under high stringency conditions. Hybridization to filter-bound sequences under low stringency conditions may, for example, be performed in 0.5 M $NaHPO_4$, 7% sodium dodecyl sulfate (SDS), 1 mM EDTA at 65°C, and washing in 0.2 x SSC/0.1 SDS at 42°C (see Ausubel et al. (eds.) 1989, Current Protocols in Molecular Biology, Vol. 1, Green Publishing Associates, Inc., and John Wiley & Sons, Inc., New York, at p. 2.10.3). Alternatively, hybridization to filter-bound sequences under high stringency conditions, may for example, be performed in 0.5 M $NaHPO_4$, 7% SDS, 1 mM EDTA at 65°C, and washing in 0.1 x SSC/0.1% SDS at 68°C (see Ausubel et al. (eds), 1989, supra). Hybridization conditions may be modified in accordance with known methods depending on the sequence of interest (see Tijssen, 1993, Laboratory Techniques in Biochemistry and Molecular Biology -- Hybridization with Nucleic Acid Probes, Part I, Chapter 2 "Overview of Principles in Hybridization and the Strategy of Nucleic Acid Probe Assays", Elsevier, New York). Generally, stringent conditions are selected to be about 5°C lower than the thermal melting point for the specific sequence at a defined ionic strength and pH.

Nucleic acids of this invention will preferably exhibit substantial identity to Cendehill, with respect to the regions of the Cendehill genome described herein which

relate to the arthrotropic phenotype of Cendehill. More preferably, such regions will have at least about 98% identity. Most preferably, there will be complete identity in the "context" of Cendehill strain-specific nucleotides or amino acids, as "context" is described herein.

With reference to nucleic acids corresponding to the first 5'NTR of Cendehill, such correspondence may be determined by predicting the folded structure of the region rather than by measuring sequence identity. Nucleic acids of this invention include a 5'NTR having a folded structure in which one or both of the terminal and medial loops is altered in size as compared to wild-type. The size of the loop may be quantified according to the number of un-paired bases in the loop region. Preferably, such alterations result in an increase in size of the loop as compared to wild-type. More preferably, such altered loops will be of at least the size of the terminal and medial loops described herein for Cendehill. Most preferably, the sequence of un-paired bases in either loop region will be substantially the same as described herein for Cendehill loops. Further, nucleic acids of this invention comprising a 5'NTR, may include a bulge which is increased in size as compared to the wild-type bulge and preferably will have at least four un-paired bases in a bulge to one side of the stem structure. Most preferably, the sequence of un-paired bases in such a bulge will be substantially as described herein for the Cendehill bulge. Determination of predicted folding of a 5'NTR is carried out as described herein using the Mfold™ 3.0 program.

Variation in the immunogenicity, yield, stability or pathogenicity of the product may readily be determined by standard techniques by comparison to known strains such as Cendehill. For example, mutation of Cendehill to increase antigenicity may be determined by measuring increased binding of a virus or viral protein to a known antibody to

rubella virus and comparing this binding to that of Cendehill virus or protein at an equivalent concentration.

Arthrotropism, for the purpose of this specification, is defined as the ability of a rubella virus strain to replicate in pieces of human joint tissue weighing approximately 0.1 gram cultured in 2 mls of medium and yield virus of titres greater than 100 plaque-forming units per ml of medium, at 24 hours post-infection that increases over the next 24 to 48 hours. Any virus less than 100 pfu/cell and that does not show an increase in titre represents residual virus from the inoculum. Following a period to allow adsorption of virus in the inoculum to the cells (4 hours), the joint pieces are washed 4 to 5 times to reduce this residual virus and characteristically 10-100 pfu/ml of virus remains after this procedure.

This invention also provides a method for constructing chimeric rubella viral strains comprising part Cendehill and part of a second rubella strain including steps whereby:

(a) cDNA for one or more of the Cendehill non-translated regions, non-structural proteins p150 and p90 and structural proteins C, E2 and E1 is joined to cDNA of a rubella virus other than Cendehill to produce DNA corresponding to a complete RNA genome of a chimeric rubella virus. This may also be incorporated into a plasmid or viral vector to provide a chimeric infectious clone.

(b) the resulting altered cDNA clone may be transcribed to produce RNA which may be used to transfect cells to produce chimeric virus, which can be cultivated as a seed stock for vaccine production.

This invention also provides rubella cDNA, RNA, or virus wherein cDNA or RNA encoding one or more of the viral

p150 or p90 proteins or the cDNA or RNA corresponding to a 5' non-translated region is derived from or is mutated to correspond to Cendehill, and at least part of the DNA, RNA or viral RNA, is derived from or is mutated to correspond to rubella other than Cendehill. Preferably, the cDNA, RNA or genome of the virus will have one or more substitutions or deletions (as compared with Therien strain) in or near the 5' non-translated region in the areas of nucleotides 17-65; substitutions in the non-structural gene coding region resulting in one or more mutations of amino acids 929, 1006, 1041, 1162 of p150 protein or amino acid 1496 of p90 protein; or, substitutions at or near nucleotides 118 or 358 of the non-structural gene encoding region.

This invention also provides the use of the aforementioned cDNA, RNA, vectors (including infectious clones) and viruses (recombinant or chimeric) in the production of modified rubella cDNA, RNA or viruses, production of modified rubella protein, and in the production of rubella vaccines (DNA vaccines, live attenuated viral vaccines and subunit vaccines).

This invention also provides the entire sequence of the Cendehill strain of rubella virus, including the identification of nucleotide substitutions relative to wild-type strains which are unique to the Cendehill strain and are associated with the attenuating phenotype. This phenotype includes temperature sensitivity and the restriction of growth in human joint tissue. These substitutions can be incorporated into other rubella strains such as the current RA27/3 vaccine to produce new vaccine strains that are not arthritogenic. Such substitutions may be in the region of nucleotides 17-65 (in or near the first 5' non-translated region) which forms a stem-loop structure. The substitutions may be at or near nucleotides 118 or 358 of the non-structural gene region, or the substitutions may involve one or more mutations of

amino acids 929, 1006, 1041, 1162 of p150 or amino acid 1496 of p90.

This invention also identifies mutations in Cendehill virus structural gene regions associated with reduced immunogenicity of this strain. These include two amino acid substitutions in the E2 protein at amino acids 306 and 413 (ie. at nucleotides 7428 or 7746/47), and four amino acid substitutions in E1 at amino acids 759, 785, 890 and 915 (ie. at nucleotides 8786/88; 8864; 9180; or 9254). Alterations of some or all of these nucleotides to the equivalent nucleotides found in a more immunogenic strain such as RA27/3 or wild-type, enables production of a modified Cendehill strain which would be more antigenic. This may also be used as an alternative vaccine.

The infectious clone of Cendehill strain exemplified herein and identified as pJCND, comprises a DNA copy of the full-length Cendehill viral genome inserted into a vector from which RNA transcripts of the genome can be synthesized *in vitro* and which transcripts are infectious when transfected into cells. In the case of pJCND, the vector is the plasmid pCL 1921, which was originally constructed by Lerner and Inouye (1990) but modified by incorporation of the pUC19 polycloning region (Yanisch-Perron et al., 1985) and an SP6 RNA polymerase promoter . This plasmid is replicated at low copy number (approximately 5 copies per cell) and contains a spectinomycin resistance gene. Transcription of pJCND or other infectious clones employing Cendehill cDNA with a suitable polymerase (eg. SP6 polymerase for pJCND) enables the production of infectious Cendehill RNA which can be transfected into cells to yield a seed stock for obtaining recombinant rubella virus stocks and rubella vaccines.

Methods for production of infectious clones, subsequent expression of RNA, transfection of cells with

such RNA and production of virus as well as use of such virus in the preparation of rubella vaccines are known, for example as described in United States Patent 5,439,814 and 5,663,065 of Frey, et al. Suitable expression vectors for rubella cDNA include those described herein as well as others known in the art such as the pSI or pCI mammalian expression systems (Promega) which incorporate the SV40 and CMV Immediate Early enhancer/promoter systems (respectively) or bacterial plasmids such as pUC19, pGEM or PBR-322 (Promega) incorporating a suitable promoter sequence such as the SP6 promoter.

Methods for production of suitable expression vectors for use in DNA vaccines are also known. For example, cDNA derived from this invention may be expressed in pSI or pCI described above or the vector could be a viral vector modified to allow expression of foreign genes. Such vectors derived from adenovirus, retrovirus, alphavirus, or vaccinia virus are frequently modified to make them non-pathogenic to the host. Such vectors expressing cDNA derived from this invention may be used directly as a DNA vaccine.

For preparation of chimeric strains according to this invention, a preferred method is to synthesize cDNA from a second rubella virus by preparing RNA from virus of the second strain using established techniques and then performing reverse transcription and PCR (polymerase chain reaction) on the isolated RNA using primers which flank the region of interest (for example, primers FI or 18 as described herein for synthesis of the Cendehill/RA27/3 chimera). The cDNA is then subjected to restriction enzyme digestion and resulting fragments are ligated into the Cendehill infectious clone which has been similarly digested to remove the same segment. Similarly, desirable portions of the Cendehill cDNA (such as the non-translated region, or non-structural genes) may be obtained by

digestion, and the resulting fragment ligated into an infectious clone of a second rubella strain which has been similarly digested.

As exemplified herein, recombinant viruses were derived from pJCND and Therien/Cendehill chimeras. These strains were compared for their ability to grow in primary human joint cells, enabling the identification of two regions associated with growth restriction in these cells, in the non-structural gene region. The identification of these regions enables the production of further recombinant virus strains which combine the phenotypic property of joint growth restriction with the immunogenicity of other rubella virus strain such as RA27/3, M33 or Therien.

Sequencing of pJCND enabled the identification of nucleotide substitutions in Cendehill which are not present in wild-type strains. The stem-loop region which includes a 5' non-translated region and extends into the non-structural open reading frame (ORF), contributes to joint growth restriction. This region has been shown to be important in viral viability and virulence in some a-viruses, including Sindbis virus and rubella virus (Niesters & Strauss, 1990, Pogue et al., 1993, Pugachev & Frey, 1998).

In the 3' subgenomic region, which includes the structural gene region, Cendehill strain contains 67 substitutions relative to the Therien strain: three in the non-translated region (NTR) upstream of the translational start site of the subgenomic RNA, two in the 3'NTR, and the remainder in the coding region. Many of the substitutions in the structural genes occur as the third base of a codon and do not affect the amino-acid composition, leaving 16 substitutions in the 1062 amino-acids comprising the structural genes (nine of which are also found in the M33 strain). The substitutions

include two substitutions in the capsid protein, two in the E2 glycoprotein and four in the E1 glycoprotein. Modifications to the Cendehill structural genes (for example, by site specific mutagenesis, linker-insertion mutagenesis or homologous recombination) to provide a strain with higher immunogenicity while retaining the attenuating characteristics of Cendehill can therefore be carried out.

Brief Description of the Drawings

Figure 1 is a schematic showing the organization of the rubella virus genome. The RNA is polyadenylated (A_n) and both the genomic and sub-genomic species are capped (CAP).

Figure 2 describes the oligonucleotide primers used for reverse transcription of Rubella virus RNA and amplification of cDNA. Identification numbers for each primer appear on the left. Viral genome positions corresponding to nucleotide positions in Appendix I for seven of the primers, appear on the right.

Figure 3 is a schematic showing four Cendehill cDNA fragments used to construct chimeric viruses and an Cendehill infectious clone, beneath a general representation of the viral genome. Restriction sites are identified and location of sites used for construction are indicated by the dotted lines. Primers used to generate each cDNA fragment are indicated by primer identification numbers (from Figure 2) at fragment termini.

Figure 4 is a schematic showing the modified polycloning site of pCLPC, which is derived from pCL1921.

Figure 5 is a schematic of a cloning strategy for production of Cendehill and Cendehill chimeric clones.

Cendehill double stranded (ds) cDNA fragments are cut using the appropriate restriction enzymes and inserted sequentially into similarly restricted regions of pROBO302.

Figure 6 is a schematic comparing pROBO302 to a full-length Cendehill clone (pJCND) and two Cendehill chimeras (pROC3) and pROC3M). Regions without cross-hatching are Therien and cross-hatched regions are Cendehill.

Figure 7 shows predicted 5' stem loop structures of rubella RNA's generated by the Mfold™ 3.0 program using the published default settings and for linear RNA. Figures 7A, 7B and 7C are for Cendehill, wild-type and RA27/3, respectively. The wild-type structure shown in Figure 7B is the same for the Therien and M33 strains and also the HPV77 vaccine.

Figure 8 is a schematic showing the non-structural gene region and the position of amino acid substitutions in the Cendehill strain relative to Therien. Bars indicate mutations described by single letter amino acid codes.

Figure 9 is a schematic showing the structural genes, glycosylation sites and the position of the amino acid substitutions in the Cendehill strain as compared to Therien, including those shared with M33 strain (unshaded bars). Solid bars indicate mutations unique to Cendehill.

Detailed Description of Embodiments of the Invention

An infectious clone comprising a cDNA copy of all of the RNA of the Cendehill strain of rubella virus was produced as described below.

Isolation of Viral RNA

Cendehill virions were obtained by pelleting supernatant virus from the medium of Vero cells infected with Cendehill virus (Rohm Pharma) for 4 hours @ 18000 rpm in a Sorval™ centrifuge. Viral RNA was isolated by extraction with acidified phenol/guanidinium isothiocyanate using Trizol™ (Gibco/BRL) according to the manufacturer's instructions. RNA was precipitated from the aqueous phase by the addition of isopropyl alcohol (1:1) and washed with 75% ethanol diluted in DEPC-treated H_2O prior to drying and resuspension in DEPC-treated ddH_20.

Reverse Transcription

Specific primers complementary to the published sequence of the Therien strain were used to initiate the first strand of DNA synthesis. The primers used were #16, 38 and 125 (Figure 2). For each reaction, the primer was mixed with viral RNA in H_2O (total volume 11μl) and heated for 3 min @ 90°C. RNA was then transcribed using 200U of Superscript II™ (Life Technologies). The standard reaction mixture contained 10mM dithiothreitol and 1mM dNTPs. The volume was brought to 100μl by addition of TE buffer and heated to 90°C to inactivate the reverse transcriptase. Enzyme, primers and excess nucleotides were removed by extraction of the mixture with phenol/chloroform/isoamyl alcohol (25:24:1, by volume), followed by precipitation at -20°C in 0.3M sodium acetate and 66% ethanol.

Thermal Cycling Amplification

After generation of the first strand of DNA by reverse transcription, double stranded cDNA was made by thermal cycling amplification with a Minicycler™ (MJ Research) using the specific primers (described in Figure 2 according to the scheme shown in Figure 3) and repeated cycles of incubation with Deep Vent™ (NEB) thermostable polymerase with 3'-5' proof-reading exonuclease activity. The

standard reaction mixture contained 400μM dNTP, 2mM $MgSO_4$, 0.5μM primer and 1 unit of polymerase. The products were resuspended in H_2O for ligation into the plasmid vector, pCLPC, a derivative of pCL1921 with the modified cloning site shown in Figure 4.

Cloning

Four cDNA fragments (as shown in Figure 3) amplified in pCLPC (see Figure 4), were sequentially cloned into the Therien infectious clone pROBO302 (Pugachev et al. 1997). The cloning strategy is outlined in Figure 5. To confirm insertion of the correct fragments, the sequence of each clone was compared with that of pROBO302 and Cendehill cDNA sequenced directly following reverse transcription and amplification.

Two chimeric strains and a full-length Cendehill clone were produced:

(i) pROC3 which contains nucleotides 5357 to 9762 of Cendehill as shown in Figure 5 and Appendix 1, (including the entire structural gene region) and nucleotides 1 to 5356 of the Therien strain (the majority of the non-structural genes and 5' non-translated region);

(ii) pROC3M which contains nucleotides 2803 to 9762 of Cendehill (see Appendix 1) and nucleotides 1-2802 of Therien; and,

(iii) pJCND which contains the entire genomic sequence of the Cendehill strain (see Appendix 1). These are shown in Figure 6.

Screening of Constructs

The constructs were screened by restriction enzyme digestion to determine that the inserts were the correct size and had the expected restriction pattern. Each clone was also screened for infectivity as follows. Small-scale plasmid preparations were carried out by standard techniques. These preparations were linearised by restriction digestion with EcoR1 at the 3' terminus of the viral sequence. Positive-polarity viral RNA was generated by transcription from the SP6 promoter and the products were transfected into BHK21 cells by electroporation. After 2 days the supernatants were transferred to Vero cells and supernatant virus was removed for plaque titration 4 days later. The 3 constructs all gave titres of progeny virus of 10^5 - 10^6/ml after three serial passages in Vero cells. The progeny viruses were designated ROC3, ROC3M and JCND.

Phenotypic Characterisation of the Recombinant Viruses

Attenuating characteristics examined included temperature sensitivity and replication in human joint cells.

(1) Temperature sensitivity: At 39°C the Cendehill strain is growth-restricted while wild-type strains grow normally. This is believed to be an attenuating characteristic as growth of Cendehill would be limited in infected patients by even mild fever induction. All three recombinant strains did not grow at 39°C indicating that they have the attenuating phenotype. Similarly, measurements of the stability of the recombinant strains on prolonged incubation at 37°C, relative to the Therien and Cendehill parental strains, showed that the infectivity of the recombinants and Cendehill decreased rapidly to 0.5% of the input (a 200 fold reduction) in 50 hours while the reduction in Therien was only 10-fold.

(2) Growth in human joint cells: Mapping of the region of the genome associated with joint cell restriction was carried out by examining the ability of the recombinant viruses to replicate after electroporation into human synovial cells cultured according to the method of Miki and Chantler (1993). The results showed that five days following electroporation, the supernatant titre of pROC3 was the same as that for pROBO302 (the Therien clone). The titre of electroporated pROC3M was 10-fold lower and no growth was seen with pJCND on transfection of 0.5 μg of RNA in each case (see Table 1). Therefore the regions of the Cendehill genome containing sequences involved in joint cell restriction include nucleotides 2803 to 5355, which are present in pROC3M but not pROC and the 5′ end of the genome, nucleotides 1 to 2803 which are specific to pJCND.

Table I

Rubella Virus Strain	Virus yield (pfu/ml)
Therien	4.0×10^4
Cendehill	1.5×10^1
pROBO302	1.9×10^3
pROC3	2.5×10^3
pROC3M	2.4×10^2
pJCND1	no virus detected
pJCND2	no virus detected

Sequence Analysis

Further definition of the nucleotide substitutions involved in attenuation was determined by sequence analysis. The entire cDNA sequence corresponding to the Cendehill genome was determined using an automated sequencing system at the NAPS unit at the University of British Columbia employing Amplitaq Dye Terminator Cycle™ sequencing reagents (ABI) and by analysing the fluorescent products spectrophotometrically. The sequence obtained is shown in Appendix 1. It was compared with the published sequences of Therien strain (Dominguez et al., 1990, later corrected in Pugachev et al., 1997), a consensus M33 sequence (Clarke et al., 1987, Zheng et al., 1989 and Pugachev, 1997) and the RA27/3 sequence (Pugachev et al. 1997). Nucleotide substitutions specific to Cendehill strain in the area of the first 5'NTR, the non-structural and structural genes, and in the 3'NTR are described in detail below, in which the nucleotide numbering is according to the whole genome shown in Appendix 1 and the amino acid numbering is according to the polyproteins as described above.

5'Non-translated Region (NTR) and Stem-Loop Region

Two substitutions as shown in Table II were identified in this area.

Table II

nucleotide 37 :	U to C
nucleotide 55 :	A to G

These substitutions are in a stem-loop region that is believed to be important in controlling viral replication and translation. Alterations in this region destabilize the stem structure and may affect binding of cellular or viral factors important in viral replication.

The stem loop structure may be predicted by computer programs intended to generate representations of folded structures. For the purposes of this specification, stem loop structures are determined by use of the Mfold™ 3.0 program from Dr. Michael Zuker, Washington University School of Medicine (see: M. Zuker, et al.; Algorithms and Thermodynamics for RNA Secondary Structure Prediction: A Practical Guide in RNA Biochemistry and Biotechnology, J. Barciszewski & B.F.C. Clark eds., NATO ASI Series, Kluwer Academic Publishers (1999); and D.H. Mathews, et al. (1999) Expanded Sequence Dependence of Thermodynamic Parameters Provides Robust Prediction of RNA Secondary Structure J. Mol. Biol. 288, 911-940). The Mfold 3.0 program may also be obtained on the Internet at: http://mfold2.wustl.edu/~mfold/cgi-bin/nph-mfold-3.0cgi. The mfold program default settings are used with the imputed RNA sequence being designated as linear.

As shown in Figure 7A, the alteration at nucleotide 37 is in the terminal loop of the stem. With reference to

Figures 7A-C, the terminal loop of Cendehill is altered as compared to the predicted terminal loop of both wild-type and RA27/3 strains. As is also shown in Figure 7, the substitution at nucleotide 55 increases the size of the bulge in the stem of Cendehill as compared to the bulges of wild-type or RA27/3. As is shown in Figure 7, the medial loop of Cendehill is altered as compared to the medial loop which appears in both wild-type and RA27/3.

Attenuation of the wild-type rubella phenotype is expected upon alterations in the nucleotide region 15-65, particularly in the regions 20-28, 33-43 and 52-60. Alterations which increase the size of the bulge such that a bulge to one side of the stem has at least four unpaired nucleotides (such as is shown in Figure 7A) is also associated with the Cendehill phenotype.

Non-structural Gene (NSG) Region

Several mutations are found between nucleotides 2800 and 4550, including 5 mutations specific to the Cendehill strain which are present in pROC3M but not in pROC and are therefore associated with a significant restriction in joint cell growth as described in Table I. These mutations are delineated in Table III:

Table III

P150 nucleotide 2829	G to A	aa 929	cys - tyr
P150 nucleotide 3060	A to G	aa 1006	asp - gly
P150 nucleotide 3164	U to C	aa 1041	tyr - his
P150 nucleotide 3528	C to U	aa 1162	ala - val
P90 nucleotide 4530	C to U	aa 1496	thr - ile

Two of the NSG mutations lie within or in proximity to a region of homology with the alphavirus NSP3 domain while the other two are in the protease domain and on either side of cys 1151 at the catalytic site. The p90 mutation is in the helicase domain.

In addition to the foregoing, there are two mutations in the NSG region shown in Table IV which do not alter the encoded amino-acid but may influence infectivity due to changes in RNA structure.

Table IV

- nucleotide 118 C to U
(This substitution may be involved in stem-loop structures at the 5'end)

- nucleotide 358 U to C
(This substitution is in the region of rubella RNA involved in binding to the capsid protein

---

Structural Gene (SG) Region

The structural genes of rubella virus are produced from a 3327 nucleotide subgenomic RNA as represented in Figure 1. It consists of a short (78 nucleotide) 5'non-translated region (NTR), the structural genes which are translated from a single open-reading frame (ORF) and a short 3'NTR . Both the 3' and 5' NTRs are capable of forming stem-loop structures, can bind host cell proteins and are believed to be important in viral replication. In the entire subgenomic RNA, 67 nucleotide substitutions were identified in Cendehill strain when compared with the Therien strain (see Appendix 1). Two are in the 5'NTR upstream of the translational start site, two in the 3'NTR and the remainder are in the coding region. Many of the

substitutions in the structural genes occur as the third base of a codon and do not affect the amino-acid composition, leaving 16 substitutions in the 1062 amino-acids comprising the structural genes, eight of which are also found in the M33 strain. The remaining 8 amino acid substitutions are not found in the HPV77/DE5 or RA27/3 vaccine strains either. The nucleotide/amino acid substitutions specific to the Cendehill strain (other than the 5'NTR substitutions) are shown in Table V(a) - (d) in which the amino acid numbering is according to the polyprotein.

Table V(a): Protein C Region

nucleotide	6611	U to C	aa 34	ser-pro
nucleotides	6770	A to G	aa 87	thr-gly
	6771	C to G	"	"

The substitution at aa34 occurs within a stretch of 28 amino-acids (28-56) believed to be important in binding of protein C to viral RNA during encapsidation. A region between amino-acids 64 and 97 has been shown to react with a monoclonal antibody, indicating that this is an antigenic region although not one of the reported major antigenic sites.

Table V(b): Protein E2 Region

nucleotide	7428	C to U	aa 306	ala-val
nucleotides	7746	C to U	aa 413	thr-ile
	7747	G to U	"	"

The alanine to valine substitution at aa306 is a conservative change but lies within the first 26 residues of protein E2, a region which has been identified as a neutralising domain. The two changes at nucleotides 7746 and 7747 result in the loss of a Asn-X-Thr glycosylation site, one of four N-linked glycosylation sites found in Therien strain. The literature is conflicting as to whether the latter substitution is present in M33.

Table V(c): Protein E1

nucleotides	8786	A to G	aa 759	asn-asp
	8788	C to U	"	"
nucleotide	8864	C to A	aa 785	leu-met
nucleotide	9180	A to U	aa 890	his-leu
nucleotide	9254	G to A	aa 915	ala-thr

The four alterations in E1 all occur in the region of the protein which is extruded into the lumen of the endoplasmic reticulum, and is therefore also exposed on the surface of the mature virion. The first substitution at amino-acid 759 alters an asparagine to an aspartic acid residue with the resulting loss of an N-linked glycosylation site, one of three in E1, all of which are believed to be utilised. None of the substitutions in E1 are in regions identified as dominant epitopes of the cell-mediated immune response, nor in regions identified by monoclonal antibodies as being associated with hemagglutination or neutralisation. However they may alter conformation-dependent epitopes associated with the humoral response affecting the immunogenicity of Cendehill strain which reacts poorly with polyclonal antisera to the Therien strain in immunoprecipitation and immunoblot assays.

Table V'd.: 3'NTR

nuclectide	9731	G to C
nuclectide	9740	C to U
nuclectide	9741	C to U

This region, like the 5'NTR is involved in RNA replication. Although the substitutions at nucleotides 9731 and 9740 are also found in the M33 strain, they may affect attenuation as M33 is a less cytopathic strain than Therien.

The substitutions identified in the structural genes of Cendehill are responsible for the lower antigenicity and immunogenicity of this strain relative to Therien, M33 or RA27/3. Using the Cendehill infectious clone, alterations to the structural genes (for example, by site-directed metagenesis) would enable the antigenicity of this strain to be repaired. This would provide a novel rubella strain with the attenuating phenotype of Cendehill, including restriction of growth in joint cells, but with the immunogenic properties of either a wild strain like Therien or the RA27/3 vaccine strain. Alternatively, a chimeric strain can be produced comprising (for example) the entire structural gene region of RA27/3 inserted into the Cendehill infectious clone. Either of these constructs would provide an improved attenuated rubella vaccine.

Production of Modified Rubella Virus Strains

Altered strains can be produced by standard recombinant DNA technology as described in many current textbooks including "Molecular Cloning: A Laboratory Manual," edited by Maniatis,T., Fritsch E.F., and Sambrook, J., (Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y. 1989) or "Current Protocols in Molecular

Biology" edited by Ausubel et al., (Wiley Interscience, 1987).

To alter specific nucleotides in the structural gene region, oligonucleotide-directed mutagenesis and gene amplification technology can be used as described by Higuchi (1989). This procedure involves synthesis of oligonucleotides specific for the region to be modified, containing the required nucleotide substitution, as well as an appropriate restriction site. This can then be used as one primer for a gene amplification reaction encompassing the region of interest. A second primer is chosen which includes a unique restriction site and which will yield a fragment of suitable size. Following amplification of the fragment which now has the requisite nucleotide substitution incorporated, the fragment is cloned into the infectious clone replacing the original sequence. In this way, mutations can be incorporated into the gene sequence either singly or sequentially until the resulting virus has the properties wanted.

Production of Chimeric Virus Strains

A cDNA clone including the entire structural gene region of a rubella stain such as RA27/3 can be made in the following steps: (i) isolation of viral RNA from high-titre virus stock, (ii) first strand cDNA synthesis using a specific primer for the 3'end, (iii) amplification of the structural gene region using primers F1 and 18 (Figure 2), (iv) digestion of the amplified fragment and also pCND with Bgl II and EcoR1, and (v) cloning of the amplified fragment into pJCND (previously separated from its digested insert).

Following the above-described scheme, a chimeric Cendehill/RA27/3 clone whose genome includes a first portion which is equivalent to the Cendehill 5" non-

translated RNA, Cendehill p150 and p90 and a second portion equivalent to the structural gene region and the 3' non-translated region of RA27/3 strain was made. This clone can be used to produce a chimeric virus that expresses the structural proteins of RA27/3 but has the determinants of arthrotropism found in the genetic structure at the 5' end and in the non-structural genes of Cendehill strain.

This construct was produced by synthesising a cDNA/PCR fragment, using RA27/3 RNA as template, equivalent to the 18-F1 fragment shown in Figure 2. This fragment was then inserted into the Cendehill infectious clone using the restriction enzymes BglII and EcoR1, in an identical manner to the synthesis of pROC3 described elsewhere in this specification. The new chimeric clone was sequenced through nucleotides 6611 and 6770/6771 as well as through nucleotides 8786/8788 and 8864 to ensure that replacement of the 18-F1 fragment had occurred. The published sequence of RA27/3 indicates that the latter strain has the same nucleotides as Therien strain at these positions (Pugachev KV, Abernathy ES and Frey TK. Archives of Virology 142 1165-1180, 1997: Genomic sequence of the RA27/3 vaccine strain of rubella virus) while Cendehill is modified in these regions as disclosed herein.

Screening of Novel Rubella Strains

Modified cDNA clones incorporated in the pCL1921 plasmid can be transcribed into complete infectious RNA from the SP6 promoter. The RNA produced can be transfected into BHK-21 cells by a variety of techniques including electroporation or use of Lipofectamine™ (Gibco/BRL). The transfected RNA is translated and replicated in the cell to yield virus with altered phenotypic properties according to the mutations introduced. In this way, seed stocks of rubella strains of this invention may be produced.

Phenotypic properties of rubella strains of this invention can be monitored for characteristics associated with attenuation and immunogenicity. For example, yield, temperature sensitivity and the ability to grow in human joint tissue can be determined as described previously for pROC3 and pROC3M. The antigenicity of the strains can be assessed using standard enzyme-linked immunosorbent assays, immunoprecipitation assays and immunoblots with human rubella seropositive antisera. The efficacy of a strain for eliciting a strong neutralising antibody response can be measured in rabbits and compared with the current vaccine strain, RA27/3 and also the parental Cendehill strain. In this way, novel strains can be assessed for characteristics that would make them suitable for use as improved attenuated vaccines.

Attenuated rubella strains may be used as a seed stock for manufacturing vaccine. Virus from such a stock may be combined with a variety of stabilisers such as saline, phosphate buffer, polyethylene glycol, glycerin as currently used in vaccine preparations. The vaccine may be produced in lyophilised form to aid long-term preservation. It can also be combined with other vaccines such as mumps and measles vaccines as in the current M-M-R formulation.

In addition to use of rubella virus strains of this invention as live attenuated vaccines as described above, modified infectious cDNA clones may also be used to produce a DNA vaccine against rubella virus, either singly or in combination with other DNA vaccines. For this, the cDNA of the rubella virus strain is sub-cloned into an expression vector (either plasmid or viral) which contains a suitable eukaryotic promoter. Either the entire rubella virus genome, the structural genes or immunogenic regions of the structural genes can be used in this manner to directly immunise patients. The DNA vaccine is taken up by cells and transcribed from the eukaryotic promoter to yield RNA

which is translated into viral proteins. These in turn elicit an immune response.

Other uses of the Cendehill infectious clone and its derivatives include the production of large quantities of virus for use as antigen in enzyme-linked immunosorbent assays to assess human antibody levels against rubella. In view of variations in the antigenicity of the different rubella virus strains, it would be preferable to use antigen known to react optimally according to the vaccine strain delivered. For example, a virus strain with the structural gene region identical to the vaccine in use, but altered in the non-structural genes or NTR regions to improve viral yield for antigen production may be propagated. Subsequently, the strain for use in immunoassays would be treated to produce a non-infectious antigen preparation. Alternatively, the structural proteins alone could be produced from a suitable expression vector to yield an antigen preparation with the correct specificity.

Although various aspects of the present invention have been described in detail, it will be apparent that changes and modification of those aspects described herein will fall within the scope of the appended claims. All publications and references referred to herein are hereby incorporated by reference.

References

Best J.M., Banatvala J.B. and Bowen J.M. (1974). New Japanese rubella vaccine: Comparative trials. *Brit. Med. J.* 3:221-224.

Clarke D.M., Loo T.W., Hui I. and Chong P. Gillam S. (1987). Nucleotide sequence and *in vitro* expression of rubella virus 24S subgenomic messenger RNA encoding the

structural proteins E1, E2 and C. *Nucl. Acids. Res.* 15:3041-3057.

Dominquez G., Wang C.Y. and Frey T.K. (1990). Sequence of the genome RNA of rubella virus: Evidence for genetic rearrangement during Togavirus evolution. *Virol.* 177:225-238.

Ford D.K., Tingle A.J. and Chantler J.K. (1988). Rubella arthritis in "*Infections in the Rheumatic Diseases*". Eds. Espinoza, Alarcon, Arnett, Goldenberg. Published by Grune & Stratton. Ch. 14:103-108.

Frey T.K. and Marr L.D. (1988). Sequence of the region coding for virion proteins C and E2 and the carboxy terminus of the nonstructural proteins of rubella virus: comparison with alphaviruses. *Gene*. 62:85-99.

Higuchi R., (1990). *PCR Technology. Principles and Applications for DNA amplifications*, Ed. H.A. Erlich, Stockton Press. Chapter 6, 61-70.

Lerner C.G. and Inouye M. (1990). Low copy number plasmids for regulated low-level expression of cloned genes in *Escherichia coli*. with blue/white insert screening capacity. *Nucl. Acids. Res.* 18-4631.

Meyer H.M. Jr., Parkman P.D. and Hobbins T.E. (1969). Attenuated rubella viruses. Amer. J. Dis. Child.

Miki N.P.H. and Chantler J.K. (1992). Differential ability of wild-type and vaccine strains of rubella virus to replicate and persist in human joint tissue. *Clin. Exp. Rheumatol.* 10:3-12.

Niesters H.G. and Strauss (1990). Mutagenesis of the conserved 51 nucleotide region of Sindbis virus. *J. Virol.* 64:1639-1647.

Peetermans J. and Huygelen C. (1967). Attenuation of rubella virus by serial passage in primary rabbit kidney cell cultures. *Arch. fur Virusforsch* 21:133-142.

Plotkin S.A. and Buser F. (1985). History of RA27/3 rubella vaccine. *Res. Inf. Dis.* 7:577-578.

Plotkin S.A., Farquhar J.D. Katz M. and Buser F. (1973). Attenuation of RA27/3 rubella virus in WI-38 human diploid cells. *Amer. J. Dis. Child.* 118:178-1.

Poque D.P., Cao X.Q., Singh N.K. and Nakhasi H.L. (1993). 5' sequences of rubella virus RNA stimulate translation of chimeric RNAs and specifically interact with two host-encoded proteins. *J. Virol.* 67:7106-7117.

Pugachev K.V., Abernathy E.S. and Frey T.K. (1997a). Genomic sequence of the RA27/3 vaccine strain of rubella virus. *Arch. Virol.* 142:1165-1180.

Pugachev K.V., Abernathy E.S. and Frey T.K. (1997b). Improvement of the specific infectivity of the rubella virus (RUB) infectious clone: determinants of cytophathogenicity induced by RUB map to the non-structural proteins. *J. Virol.* 71:562-568.

Pugachev K.V. and Frey T.K. (1998). Effects of defined mutations in the 5'NTR of rubella virus genomic RNA on virus viability and macromolecular synthesis. *J. Virol.* 72:641-650.

Yanisch-Perron C., Vleira J. and Messing J. (1985). Improved M13 phage cloning vectors and host strains:

nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene*. 33:103-119.

Zheng D., Dickens L., Liu T.Y. and Nakhasi H.L. (1989). Nucleotide sequence of the 24S subgenomic messenger RNA of a vaccine strain (HPV77) of rubella virus: comparison with a wild-type strain (M33). *Gene*. 82:343-349.

APPENDIX 1

Sequence of Cendehill virus cDNA

```
                5' NTR                                                  →p150
1       CAATGGAAGC TATCGGACCT CGCTTAGGAC TCCTATCCCC ATG GAG AAA

50      CTC CTG GAT GAG GTT CTT GCC CCC GGT GGG CCT TAT AAC TTA ACC GTC GGC
101     AGT TGG GTA AGA GAC CAT GTC CGC TCA ATT GTC GAG GGC GCG TGG GAA GTG
152     CGC GAT GTT GTT ACC GCT GCC CAA AAG CGG GCC ATC GTA GCC GTG ATA
200     CCC AGA CCT GTG TTC ACG CAG ATG CAG GTC AGT GAT CAC CCA GCA CTC CAC
251     GCA ATT TCG CGG TAT ACC CGC CGC CAT TGG ATC GAG TGG GGC CCT AAA GAA
302     GCC CTA CAC GTC CTC ATC GAC CCA AGC CCG GGC CTG CTC CGC GAG GTC
350     GCT CGC GTC GAG CGC CGC TGG GTC GCA CTG TGC CTC CAC AGG ACG GCA CGC
401     AAA CTC GCC ACC GCC CTG GCC GAG ACG GCC AGC GAG GCG TGG CAC GCT GAC
451     TAC GTG TGC GCG CTG CGT GGC GCA CCG AGC GGC CCC TTC TAC GTC CAC
502     CCC GAG GAC GTC CCG CAC GGC GGT CGC GCC GTG GCG GAC AGA TGC TTG CTC
551     TAC TAC ACA CCC ATG CAG ATG TGC GAG CTG ATG CGC ACC ATT GAC GCC ACC
602     TTG CTC GTG GCG GTT GAC TTG TGG CCG GTC GCC CTT GCG GCC CAC GTC
650     GGC GAT GAC TGG GAC GAC CTG GGC ATT GCC TGG CAT CTC GAC CAT GAC GGC
701     GGT TGC CCC GCC GAT TGT CGT GGA GCC GGC GCT GGG CCC ACG CCC GGC TAC
752     ACC CGC CCC TGC ACC ACA CGC ATC TAC CAA GTC CTG CCG GAC ACC GCC
800     CAC CCC GGG CGC CTC TAC CGG TGC GGG CCC CGC CTG TGG ACG CGC GAT TGC
851     GCC GTG GCC GAA CTC TCA TGG GAG GTT GCC CAA CAC TGC GGG CAC CAG GCG
902     CGC GTG CGC GCC GTG CGA TGC ACC CTC CCT ATC CGC CAC GTG CGC AGC
950     CTC CAA CCC AGC GCG CGG GTC CGA CTC CCG GAC CTC GTC CAT CTC GCC GAA
1001    GTG GGC CGG TGG CGG TGG TTC AGC CTC CCC CGC CCC GTG TTC CAG CGC ATG
1052    CTG TCC TAC TGC AAG ACC CTG AGC CCC GAC GCG TAC TAC AGC GAG CGC
1100    GTG TTC AAG TTC AAG AAC GCC CTG AGC CAC AGC ATC ACG CTC GCG GGC AAT
1151    GTG CTG CAA GAG GGG TGG AAG GGC ACG TGC GCC GAA GAA GAC GCG CTG TGC
1202    GCA TAC GTA GCC TTC CGC GCG TGG CAG TCT AAC GCC AGG TTG GCG GGG ATT
1253    ATG AAA GGC GCG AAG CGC TGC GCC GCC GAC TCT TTG AGC GTG GCC GGC TGG
1304    CTG GAC ACC ATT TGG GAC GCC ATT AAG CGG TTC TTC GGT AGC GTG CCC CTC
1355    GCC GAG CGC ATG GAG GAG TGG GAA CAG GAC GCC GCG GTC GCC GCC TTC
1403    GAC CGC GGC CCC CTC GAG GAC GGC GGG CGC CAC TTG GAC ACC GTG CAA CCC
1454    CCA AAA TCG CCG CCC CGC CCT GAG ATC GCC GCG ACC TGG ATC GTC CAC GCA
```

1505 GCC AGC GCA GAC CGC CAT TGC GCG TGC GCT CCC CGC TGC GAC GTC CCG

1553 CGC GAA CGT CCT TCC GCG CCT GCC GGC CCG CCG GAT GAC GAG GCG CTC ATC

1604 CCG CCG TGG CTG TTC GCC GAG CGC CGT GCC CTC CGC TGC CGC GAG TGG GAT

1655 TTC GAG GCT CTC CGC GCG CGC GCC GAT ACG GCG GCC GCG CCC GCC CCG

1703 CTG GCT CCA CGC CCT GCG CGG TAC CCC ACC GTG CTC TAC CGC CAC CCC GCC

1754 CAC CAC GGT CCG TGG CTC ACC CTT GAC GAG CCG GGC GAG GCT GAC GCG GCC

1805 CTG GTC TTA TGC GAC CCA TTT GGC CAG CCG CTC CGG GGC CCT GAA CGC

1853 CAC TTC GCC GCC GGC GCG CAT ATG TGC GCG CAG GCG CGG GGG CTC CAG GCT

1904 TTT GTC CGT GTC GTG CCT CCA CCC GAG CGC CCC TGG GCT GAC GGG GGC GCC

1955 AGA GCG TGG GCG AAG TTC TTC CGC GGC TGC GCC TGG GCG CAG CGC TTG

2003 CTC GGC GAG CCG GCA GTC ATG CAC CTC CCA TAC ACC GAT GGC GAC GTG CCA

2054 CAG CTG ATC GCA CTG GCC TTG CGC ACG CTG GCC CAA CAG GGG GCC GCC TTG

2105 GCA CTC TCG GTG CGT GAC CTG CCC GGG GGT GCA GCG TTC GAC GCA AAT

2153 GCG GTC ACC GCC GCC GTG CGC GCT GGC CCC GGC CAG CTC GCG GCC ACG TCA

2204 CCG CCA CCC GGC GAC CCC CCG CCG CCG CGC CGC GCA CGG CGA TCG CAA CGG

2255 CAC TCG GAC GCC CGC GGC ACT CCG CCC CCC GCG CCT GTG CGC GAC CCG

2303 CCG CCG CCC GCC CCC AGC CCG CCC GCG CCA CCC CGC GCG GGT GAC CCG GTC

2354 CCT CCC ACT CCC GCG GAG CCG GCG GAT CGC GCG CGT GAC GCC GAG CTG GAG

2405 GTC GCC TAC GAA CCG AGC GGC CCC CCC ACG TCA ACC AAG GCA GAC CCG

2453 GAC AGC GAC ATC GTT GAA AGT TAC GCC CGC GCC GCC GGA CCT GTG CAC CTC

2504 CGA GTC CGC GAC ATC ATG GAC CCA CCG CCT GGC TGC AAG GTT GTG GTC AAC

2555 GCC GCC AAC GAG GGG CTG CTG GCC GGC TCC GGC GTG TGC GGT GCC ATC

2603 TTT GCC AAC GCC ACG GCG GCC CTC GCT GCA GAC TGC CGG CGC CTC GCC CCA

2654 TGC CCC ACC GGC GAG GCG GTG GCG ACA CCC GGC CAC GGC TGC GGG TAC ACC

2705 CAC ATC ATC CAC GCC GTC GCG CCG CGG CGT CCT CGG GAC CCC GCC GCC

2753 CTC GAG GAG GGC GAA GCG CTG CTC GAG CGC GCC TAC CGC AGC ATC GTC GCG

2804 CTA GCC GCC GCG CGT CGG TGG GCG TAT GTC GCG TGC CCC CTC CTC GGC GCT

2855 GGC GTC TAC GGC TGG TCT GCT GCG GAG TCC CTT CGA GCC GCG CTC GCG

2903 GCT ACG CGC GCC GAG CCC GTC GAG CGC GTG AGC CTG CAC ATC TGC CAC CCC

2954 GAC CGC GCC ACG CTG ACG CAC GCC TCC GTG CTC GTC GGC GCG GGG CTC GCT

3005 GCC AGG CGC GTC AGT CCT CCT CCG ACC GAG CCC CTC GCA TCT TGC CCC

3053 GCC GGT GGC CCG GGC CGA CCG GCT CAG CGC AGC GCG TCG CCC CCA GCG ACC

3104 CCC CTT GGG GAT GCC ACC GCG CCC GAG CCC CGC GGA TGC CAG GGG TGC GAA

3155 CTC TGC CGG CAC ACG CGC GTC ACC AAT GAC CGC GCC TAC GTC AAC CTG

3203 TGG CTC GAG CGC GAC CGC GGC GCC ACC AGC TGG GCG ATG CGC ATT CCC GAG

3254 GTG GTT GTC TAC GGG CCG GAG CAC CTC GCC ACG CAT TTT CCA CTA AAC CAC

3305 TAC AGT GTG CTC AAG CCC GCG GAG GTC AGG CCC CCG CGA GGC ATG TGC

3353 GGG AGT GAC ATG TGG CGC TGC CGC GGC TGG CAG GGC ATG CCG CAG GTG CGG

3404 TGC ACC CCC TCC AAC GCT CAC GCC GCC CTG TGC CGC ACA GGC GTG CCC CCT

3455 CGG GTG AGC ACG CGA GGC GGC GAG CTA GAC CCA AAC ACC TGC TGG CTC

3503 CGC GCC GCC GCC AAC GTT GCG CAG GTT GCG CGC GCC TGC GGC GCC TAC ACG

3554 AGT GCC GGG TGC CCC AAG TGC GCC TAC GGC CGC GCC CTG AGC GAA GCC CGC

3605 ACT CAT GAG GAC TTC GCC GCG CTG AGC CAG CGG TGG AGC GCG AGC CAC

3653 GCC GAT GCC TCC CCT GAC GGC ACC GGA GAC CCC CTC GAC CCC CTG ATG GAG

3704 ACC GTG GGA TGC GCC TGT TCG CGC GTA TGG GTC GGC TCC GAG CAA GAG GCC

3755 CCG CCC GAC CAC CTC CTG GTG TCC CTC CAC CGT GCC CCC AAT GGT CCG

3803 TGG GGC GTA GTG CTC GAG GTG CGT GCG CGC CCC GAG GGG GGC AAC CCC ACC

3854 GGC CAC TTC GTC TGC GCG GTC GGC GGC GGC CCA CGC CGC GTC TCG GAC CGC

3905 CCC CAC CTT TGG CTC GCG GTC CCC CTG TCT CGG GGC GGC GGC ACT TGT

p90

3953 GCC GCG ACC GAC GAG GGG CTG GCC CAG GCG TAC TAC GAC GAC CTC GAG GTG

4004 CGC CGC CTC GGG GAT GAC GCC ATG GCC CGG GCG GCC CTC GCA TCA GTC CAA

4055 CGC CCT CGC AAA GGC CCT TAC AAT ATC AGG GTA TGG AAC ATG GCC GCA

4103 GGC GCT GGC AAG ACC ACC CGC ATC CTC GCT GCC TTC ACG CGC GAA GAC CTC

4154 TAC GTC TGC CCC ACC AAT GCG CTC CTG CAC GAG ATC CAG GCC AAA CTC CGC

4205 GCG CGC GAT ATC GAC ATC AAG AAC GCC GCC ACC TAC GAG CGC GCG CTG

4253 ACG AAA CCG CTC GCC GCC TAC CGA CGC ATC TAC ATC GAT GAG GCG TTC ACT

4304 CTC GGC GGC GAG TAC TGC GCG TTC GTT GCC AGC CAA ACC ACC GCG GAG GTG

4355 ATC TGC GTC GGT GAT CGG GAC CAG TGC GGC CCA CAC TAC GCC AAT AAC

4403 TGC CGC ACC CCC GTC CCT GAC CGC TGG CCT ACC GAG CGC TCA CGC CAC ACT

4454 TGG CGC TTC CCC GAC TGC TGG GCG GCC CGC CTG CGC GCG GGG CTC GAT TAT

4505 GAC ATC GAG GGC GAG CGC ACC GGC ATC TTC GCC TGC AAC CTT TGG GAC

4553 GGC CGC CAG GTC GAC CTT CAC CTC GCC TTC TCG CGC GAA ACC GTG CGC CGC

4604 CTT CAC GAG GCT GGC ATA CGC GCA TAC ACC GTG CGC GAG GCC CAG GGT ATG

4655 AGC GTC GGC ACC GCC TGC ATC CAT GTG GGC AGG GAC GGC ACG GAC GTT

4703 GCC CTG GCG TTG ACA CGC GAC CTC GCC ATC GTC AGC CTG ACC CGG GCC TCC

4754 GAC GCC CTC TAC CTC CAC GAG CTC GAG GAC GGC TCA CTG CGC GCT GCG GGG

4805 CTC AGC GCG TTC CTC GAC GCT GGG GCA CTG GCG GAA CTC AAG GAG GTT

4853 CCC GCC GGC ATT GAC CGC GTT GTC GCC GTC GAG CAG GCA CCA CCA CCG TTG

4904 CCG CCC GCC GAC GGC ATC CCC GAG GCC CAA GAC GTG CCG CCC TTC TGC CCC

4955 CGC ACT CTA GAG GAG CTC GTC TTC GGC CGT GCC GGC CAC CCC CAT TAC

5003 GCG GAC CTC AAC CGC GTG ACT GAG GGC GAA CGA GAA GTG CGG TAT ATG CGC

5054 ATC TCG CGT CAC CTG CTC AAC AAG AAT CAC ACC GAG ATG CCC GGA ACG GAA

5105 CGC GTT CTC AGT GCC GTT TGC GCC GTG CGG CGC TAC CGC GCG GGC GAG GAT

5156 GGG TCG ACC CTC CGC ACT GCT GTG GCC CGC CAG CAC CCG CGC CCT TTC CGC

5207 CAG ATC CCA CCC CCG CGC GTC ACT GCT GGG GTC GCC CAG GAG TGG CGC ATG

5258 ACG TAC TTG CGG GAA CGG ATC GAC CTC ACT GAT GTC TAC ACG CAG ATG

5306 GGC GTG GCC GCG CGG GAG CTC ACC GAC CGC TAC GCG CGC CGC TAT CCT GAG

5357 ATC TTT GCC GGC ATG TGT ACC GCC CAG AGC CTG AGC GTC CCC GCC TTC CTC

5408 AAA GCC ACC TTG AAG TGC GTA GAC GCC GCC CTC GGC CCC AGG GAC ACC

5456 GAG GAC TGC CAC GCC GCT CAG GGG AAA GCC GGC CTT GAG ATT CGG GCA TGG

5507 GCC AAG GAG TGG GTC CAG GTT ATG TCC CCG CAT TTC CGC GCG ATC CAG AAG

5558 ATT ATC ATG CGC GCC TTG CGC CCG CAA TTC CTT GTG GCC GCT GGC CAT

5606 ACG GAG CCC GAG GTC GAT GCG TGG TGG CAG GCT CAT TAC ACC ACC AAC GCC

5657 ATC GAG GTC GAC TTC ACT GAG TTC GAC ATG AAC CAG ACC CTC GCT ACT CGG

5708 GAC GTC GAG CTC GAG ATT AGC GCC GCT CTC TTG GGC CTC CCT TGC GCC

5756 GAA GAC TAC CGC GCG CTC CGC GCC GGC AGC TAT TGC ACT CTG CGC GAA CTG

5807 GGC TCC ACT GAG ACC GGC TGC GAG CGC ACA AGC GGC GAG CCT GCC ACG CTG

5858 CTG CAC AAC ACC ACC GTG GCC ATG TGC ATG GCT ATG CGC ATG GTC CCC

5906 AAA GGC GTG CGC TGG GCC GGG ATT TTC CAG GGC GAC GAT ATG GTC ATC TTC

5957 CTC CCC GAG GGC GCG CGC AGT GCG GCA CTC AAG TGG ACC CCC GCC GAG GTG

6008 GGC TTG TTC GGC TTC CAC ATT CCA GTG AAG CAT GTG AGC ACC CCA ACC

6056 CCC AGC TTC TGC GGG CAC GTC GGC ACC GCG GCC GGC CTC TTC CAT GAT GTC

6107 ATG CAC CAG GCA ATC AAG GTG CTT TGC CGC CGT TTC GAC CCC GAC GTG CTT

6158 GAA GAA CAG CAG GTG GCC CTC CTC GAC CGC CTC CGG GGG GTC TAC GCG

6206 GCT CTG CCT GAC ACC GTT GCC GCC AAT GCT GCG TAC TAT GAC TAC AGC GCG

6257 GAG CGC GTC CTC GCT ATC GTG CGC GAA CTT ACC GCG TAC GCG CGG GGG CGC

6308 GGC CTC GAC CAC CCG GCC ACC ATC GGC GCG CTC GAG GAG ATT CAG ACC

6356 CCC TAC GCG CGC GCC AAT CTC CAC GAC GCT GAC TAA CGC CCC CGT ACG TGG

NTR

```
                                                    ┌→subgenome (NTR)
6401  GGC CTT TAA TCT CAC CTA CTC TAA CCA │GGTCATCACC CACCGTTGTT

6451  TCGCCGCATC TGGTGGGTAC CCCACTCTTG CCATTCGGGA GAGCCCCAGG GTGCCCGA

                 >C
6501  ATG GCT TCC ACT ACC CCC ATC ACC ATG GAG GAC CTT CAG AAG GCC

6551  CTC GAG GCA CAA TCC CGC GCC CTG CGC GCG GAA CTC GCC GCC GGC GCC TCG

6601  CAG CCG CGC CGG CCG CGG CCG CCG CGA CAG CGC GAC TCC AGC ACC TCC

6651  GGA GAT GAC TCC GGC CGT GAC TCC GGA GGG CCC CGC CGC CGC CGC GGC AAC

6701  CGG GGC CGT GGC CAG CGC AAG GAC TGG TCC AGG GCC CCG CCC CCC CCG GAA

6751  GAG CGG CAA GAA GGT CGC TCC CAA ACT CCG GCC CCG AAG CCA TCG CGG

6801  GCG CCG CCA CAA CAG CCT CAA CCC CCG CGC ATG CAA ACC GGG CGT GGG GGT

6851  TCT GCC CCG CGC CCT GAG CTG GGG CCG CCG ACC AAC CCG TTC CAG GCA GCC

6901  GTG GCG CGT GGC CTC CGC CCG CCT CTC CAT GAC CCT GAT ACC GAG GCA

6951  CCC ACC GAG GCC TGC GTG ACC TCA TGG CTT TGG AGC GAG GGC GAA GGC GCG

7001  GTC TTC TAC CGC GTC GAC CTG CAT TTC ACC AAC CTG GGC ACC CCC CCA CTC

7051  GAC GAG GAC GGC CGC TGG GAC CCT GCG CTC ATG TAC AAC CCT TGC GGG

7101  CCT GAG CCG CCT GCT CAC GTC GTC CGC GCG TAC AAC CAA CCT GCC GGC GAC

7151  GTC AGG GGC GTT TGG GGT AAA GGC GAG CGC ACC TAC GCC GAG CAG GAT TTC

7201  CGC GTC GGC GGC ACG CGC TGG CAC CGC CTG CTG CGC ATG CCA GTG CGC

7251  GGC CTC GAC GGC GAC AGC GCC CCG CTT CCC CCC CAC ACC ACC GAG CGC ATT

7301  GAG ACC CGC TCG GCG CGC CAT CCT TGG CGC ATC CGC TTC GGT GCC CCC CAG

7351  GCC TTC CTC GCC GGG CTC TTG CTC GCC GCG GTC GCC GTT GGC ACC GCG

                    >E2
7401  CGC GCC│GGG CTC CAG CCC CGC GTT GAT ATG GCG GCA CCC CCT ACG CCG CCG

7451  CAG CCC CCC CGT GCG CAC GGG CAG CAT TAC GGT CAC CAC CAC CAT CAG CTG

7501  CCG TTC CTC GGG CAC GAC GGC CAT CAC GGC GGC ACC TTG CGC GTC GGC

7551  CAG CAT CAC CGA AAC GCC AGC GAC GTG CTG CCC GGC CAC TGG CTC CAA GGC

7601  GGC TGG GGT TGC TAC AAC CTG AGC GAC TGG CAC CAG GGC ACT CAT GTC TGT

7651  CAC ACC AAG CAC ATG GAC TTC TGG TGT GTG GAG CAC GAC CGA CCG CCG

7701  CCC GCG ACC CCG ACG CCT CTC ACC ACC GCG GCG AAC TCC ATT ACC GCC GCC

7751  ACC CCC GCC ACT GCG CCG GCC CCC TGC CAC GCC GGC CTC AAT GAC AGC TGC

7801  GGC GGC TTC TTG TCT GGG TGC GGG CCG ATG CGC CTG CGC CAC GGC GCT

7851  GAC ACC CGG TGC GGT CGG TTG ATC TGC GGG CTG TCC ACC ACC GCC CAG TAC

7901  CCG CCT ACC CGG TTT GGC TGC GCT ATG CGG TGG GGC CTC CCC CCC TGG GAA
```

7958 CTG GTC GTT CTT ACC GCC CGC CCC GAA GAC GGC TGG ACT TGC CGC GGC

8006 GTG CCC GCC CAT CCA GGT ACC CGC TGC CCC GAA CTG GTG AGC CCC ATG GGA

8057 CGC GCG ACT TGC TCC CCA GCC TCG GCC CTC TGG CTC GCC ACA GCG AAC GCG

8108 CTG TCT CTT GAC CAC GCG CTC GCG GCC TTT GTC CTG CTG GTC CCG TGG

8156 GTC CTG ATA TTT ATG GTG TGC CGC CGC GCC TGT CGC CGC CGC GGC GCC GCC

8207 GCC GCC CTC ACC GCA GTC GTC CTG CAG GGG TAC AAC CCC CCC GCC TAT GGC

→E1

8258 GAG GAG GCT TTC ACC TAC CTC TGC ACT GCA CCG GGG TGC GCC ACT CAA

8306 ACA CCT GTC CCC GTG CGC CTC GCT GGC GTC CGC TTT GAG TCC AAG ATC GTG

8357 GAC GGC GGC TGC TTT GCC CCA TGG GAC CTC GAG GCC ACT GGA GCC TGC ATC

8408 TGC GAG ATC CCC ACT GAT GTC TCG TGC GAG GGC TTG GGG GCC TGG GTA

8456 CCC ACA GCC CCT TGC GCG CGC ATC TGG AAT GGC ACA CAG CGC GCG TGC ACC

8507 TTC TGG GCT GTC AAC GCC TAC TCC TCT GGC GGG TAC GCG CAG CTG GCC TCT

8558 TAC TTC AAC CCT GGC GGC AGC TAC TAC AAG CAG TAC CAC CCC ACC GCG

8606 TGC GAG GTT GAA CCT GCC TTC GGA CAC AGC GAC GCG GCC TGC TGG GGC TTC

8657 CCC ACC GAC ACC GTG ATG AGC GTG TTC GCC CTT GCT AGC TAC GTC CAG CAC

8708 CCT CAC AAG ACC GTC CGG GTC AAG TTT CAT ACA GAG ACT AGG ACC GTC

8756 TGG CAA CTC TCC GTA GCC GGC GTG TCG TGC GAT GTC ACC ACT GAA CAC CCG

8807 TTC TGC AAC ACG CCG CAC GGA CAA CTC GAG GTC CAG GTC CCG CCC GAC CCT

8858 GGG GAC ATG GTT GAG TAC ATT ATG AAT TAC ACC GGC AAT CAA CAG TCC

8906 CGG TGG GGC CTC GGG AGC CCG AAC TGT CAT GGC CCC GAT TGG GCC TCC CCG

8957 GTT TGC CAA CGC CAT TCC CCT GAC TGC TCG CGG CTT GTG GGG GCC ACG CCA

9008 GAG CGT CCC CGG CTG CGC CTG GTC GAC GCC GAC GAC CCC CTG CTG CGC

9056 ACT GCC CCT GGG CCC GGC GAG GTG TGG GTC ACG CCT GTC ATA GGC TCT CAG

9107 GCG CGC AAG TGC GGA CTC CAC ATA CGC GCT GGA CCG TAC GGC CAT GCT ACC

9158 GTC GAA ATG CCC GAG TGG ATC CTC GCC CAC ACC ACT AGC GAC CCC TGG

9206 CAC CCA CCG GGC CCC TTG GGG CTG AAG TTC AAG ACA GTT CGC CCG GTG ACC

9257 CTG CCA CGC GCG TTA GCG CCA CCC CGC AAT GTG CGT GTG ACC GGT TGC TAC

9308 CAG TGC GGT ACC CCC GCG CTG GTG GAA GGC CTT GCC CCA GGG GGA GGG

9356 AAC TGC CAT CTC ACC GTC AAT GGC GAG GAC GTC GGC GCC TTC CCC CCT GGG

9407 AAG TTC GTC ACC GCC GCC CTC CTC AAC ACC CCC CCG CCC TAC CAA GTC AGC

9458 TGC GGG GGC GAG AGC GAT CGC GCG AGC GCG CGG GTC ATT GAC CCC GCC

9506 GCG CAA TCG TTT ACC GGC GTG GTG TAT GGC ACA CAC ACC ACT GCT GTG TCG

9557 GAG ACC CGG CAG ACC TGG GCG GAG TGG GCT GCT GCT CAT TGG TGG CAG CTC

9608 ACT CTG GGC GCC ATT TGC GCC CTC CCA CTC GCT GGC TTA CTC GCT TGC **3' NTR**

9656 TGT GCC AAA TGC TTG TAC TAC TTG CGC GGC GCT ATA GCG CCG CGC TAG TGG

9707 GCCCCCGCGC GAAACCCGCA CTAGCCCACT AGATTTCCGC ACCTGTTGCT GTATAG

WE CLAIM:

1. A nucleic acid corresponding to a nucleic acid encoding a Cendehill rubella protein selected from the group consisting of: p150; p90; C; E1; E2.

2. A nucleic acid corresponding to a non-translated region of the Cendehill genome.

3. The nucleic acid of claim 2 wherein the non-translated region is a 5' non-translated region in which at least one of a terminal loop or a medial loop is different in size as compared to wild-type rubella 5' non-translated region.

4. A nucleic acid which includes a sequence or sequences of nucleotides corresponding to a 5' non-translated region, p90 and p150 of Cendehill.

5. DNA including a sequence of nucleotides corresponding to the entire Cendehill genome as shown in Appendix 1.

6. The nucleic acid of any one of claims 1-4 which is DNA.

7. The nucleic acid of any one of claims 1-4 which is RNA.

8. The nucleic acid of any one of claims 1-7 further including one or more sequences of nucleotides corresponding to all or part of a genome of a rubella strain other than Cendehill.

9. A plasmid or viral vector that includes a nucleic acid according to any one of claims 1-5 or 8, wherein the nucleic acid is DNA.

10. DNA comprising a sequence of nucleotides complementary to rubella genomic RNA capable of encoding an infectious virus of the Cendehill strain or having an attenuating phenotype comparable to Cendehill.

11. DNA including a first sequence of nucleotides corresponding to one or more of: a non-translated region, p150, p90, C, E1 and E2 of Cendehill strain; and, a second sequence of nucleotides that is derived from a rubella virus strain other than Cendehill, wherein said DNA encodes an infectious rubella virus.

12. DNA comprising sequences of nucleotides corresponding to nuclectides 1 to 5355 of Cendehill and nucleotides 5356 to 9762 of RA27/3.

13. The DNA of claim 10, 11 or 12, in a plasmid or viral vector capable of replication and transcription of the DNA.

14. DNA comprising one or more sequences of nucleotides encoding all or part of one or more of: p150, p90, C, E2 and E1 of Cendehill virus, incorporated into an expression vector.

15. A method of producing rubella virus comprising the steps of transcribing the DNA of claim 14 into RNA; transfecting cells with said RNA; and, recovering rubella virus from the transfected cells.

16. Rubella virus obtained by the method of claim 15 wherein the DNA transcribed includes a sequence of nucleotides derived from a rubella virus strain other than Cendehill.

17. A method of producing DNA encoding a recombinant or chimeric rubella virus exhibiting the lack of

arthrotropicity of Cendehill virus, comprising a step whereby:

(a) nuclectides in Cendehill cDNA encoding viral structural protein are altered such that the protein so encoded increases immunogenicity of a recombinant rubella virus comprising said protein;

(b) nuclectides in the non-translated regions or non-structural protein region of cDNA for rubella virus other then Cendehill are altered to decrease arthritogenicity of a recombinant rubella virus coded for by the altered cDNA; or,

(c) cDNA for one or more of a Cendehill non-translated region, non-structural protein p150, and non-structural protein p90 is joined to cDNA for a rubella virus other then Cendehill to produce DNA corresponding to a complete RNA genome of a chimeric rubella virus.

18. An infectious clone for a rubella virus comprising a vector which includes cDNA corresponding to one or more portions of Cendehill genome selected from the group consisting of: a non-translated region, protein p150, protein p90, protein C, protein E1 and protein E2; and wherein at least a part of cDNA in the infectious clone is cDNA for a rubella virus other than Cendehill.

19. A method of producing rubella RNA comprising the step of transcribing the infectious clone of claim 18.

20. Rubella RNA produced according to the method of claim 19.

21. A method of producing a rubella virus comprising the steps of transfecting cells with RNA produced according to

claim 19, and recovering rubella virus from the transfected cells.

22. A rubella virus comprising a genome including a first portion which is equivalent to one or more ribonucleic acids selected from the group consisting of: Cendehill non-translated RNA; Cendehill p150 RNA; p90 RNA; C RNA; E1 RNA; E2 RNA; and wherein a second portion of the genome is equivalent to RNA of a rubella virus other than Cendehill.

23. The virus of claim 22 wherein the virus other than Cendehill is RA27/3.

24. The virus of claim 19 or 20 wherein the first portion is all of the Cendehill 5' non-translated RNA, p150 RNA, and p90 RNA.

25. A Cendehill viral protein free of virus, selected from the group consisting of: p150, p90, C, E1 and E2, produced by expressing Cendehill cDNA encoding said protein from an expression vector.

26. Rubella cDNA, RNA, or a rubella virus having one or more nucleotide substitutions selected from the group consisting of: 37-C; 55-G; 118-T(or)U; 358-C; 2829-A; 3060-G; 3164-C; 3528-T(or)U; 4530-T(or)U; 6611-C; 6770-G; 6771-G; 7428-T(or)U; 8786-G; 8788-T(or)U; 8864-A; 9180-T(or)U; 9254-A; and 9741-T(or)U, wherein the aforesaid numbering of the nucleotide substitution is with reference to Appendix 1, and wherein said substitutions occur in the same context as shown in Appendix 1.

27. A rubella cDNA, RNA or viral genome that encodes a rubella protein selected from the group of proteins consisting of: p150/929/tyr; p150/1006/gly; p150/1041/his; p150/1162/val; p90/1496/ile; C4/pro; C/87/gly; E2/306/val;

E2/413/ile; E1/759/asp; E1/785/met; E1/890/leu; and, E1/915/thr, wherein the aforesaid proteins are identified by reference to a strain-specific amino acid in Cendehill polyprotein and wherein the strain-specific amino acid occurs in the same context as in the Cendehill polyprotein.

28. Use of DNA incorporated into an expression vector according to claim 14 as a sub-unit vaccine.

29. Use of DNA of claim 18 as a DNA vaccine.

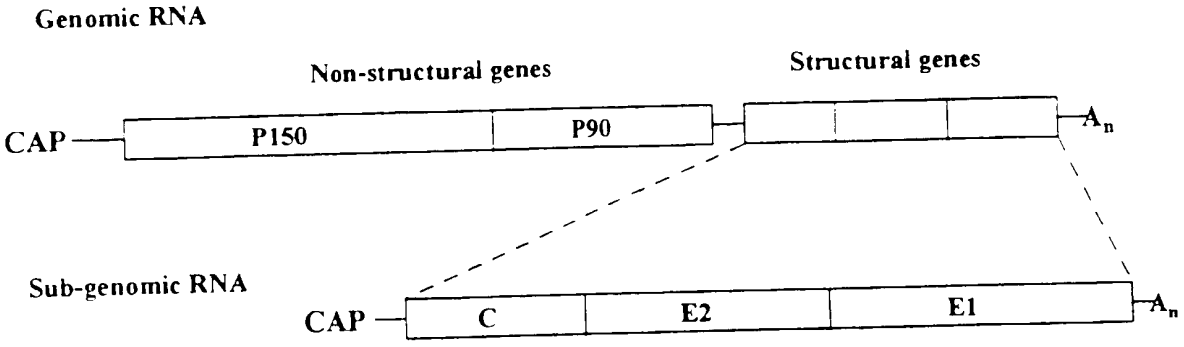


Figure 1

RV 3' end complement

F1 5'-CGCGAATTC TTTTTTTTTTTTTTTTTTTTTTTTTCTATACAGCAACAGGT
EcoR1

RV 5' start

F2 5'-TCGAAGCTTATTTAGGTGACACTATACAATGGAAGCTATCGGACCTCGCTTAGG-
Hind111 SP6

9	5'TGCAGCGTTCGACGCAAACG-	2133-2153
10	5'-TCCGAGTGCCGTTGCGATC-	2243-2262
16	5'-GCGTTCTTGATGTCGATATCGCG-	4410-4431
18	5'-CTCACTGATGTCTACACGCAGATG-	5281-5763
46	5'-CAACCACCTCGGGAATGC-	3241-3260
125	5'-TAGTCTTCGGCGCTTGG-	5747-5763
251	5'-TTTGCCAACGCCACGGC-	2603-2618

Figure 2

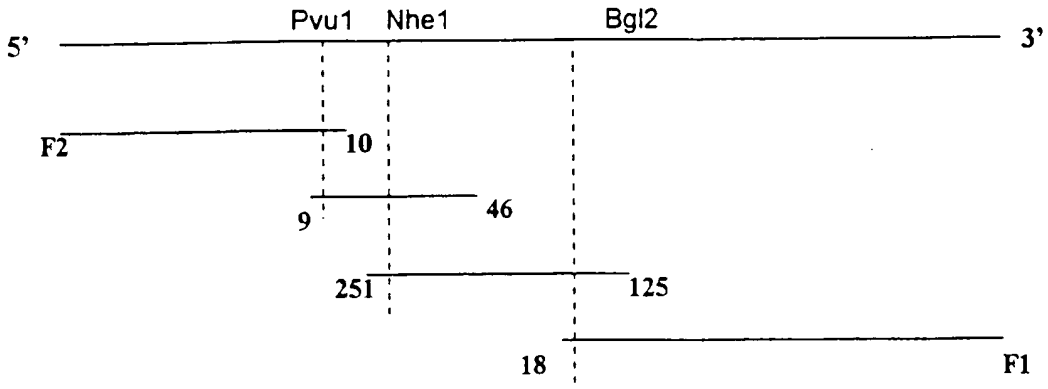


Figure 3

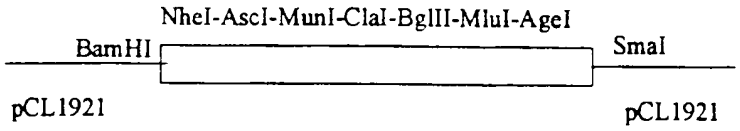


Figure 4

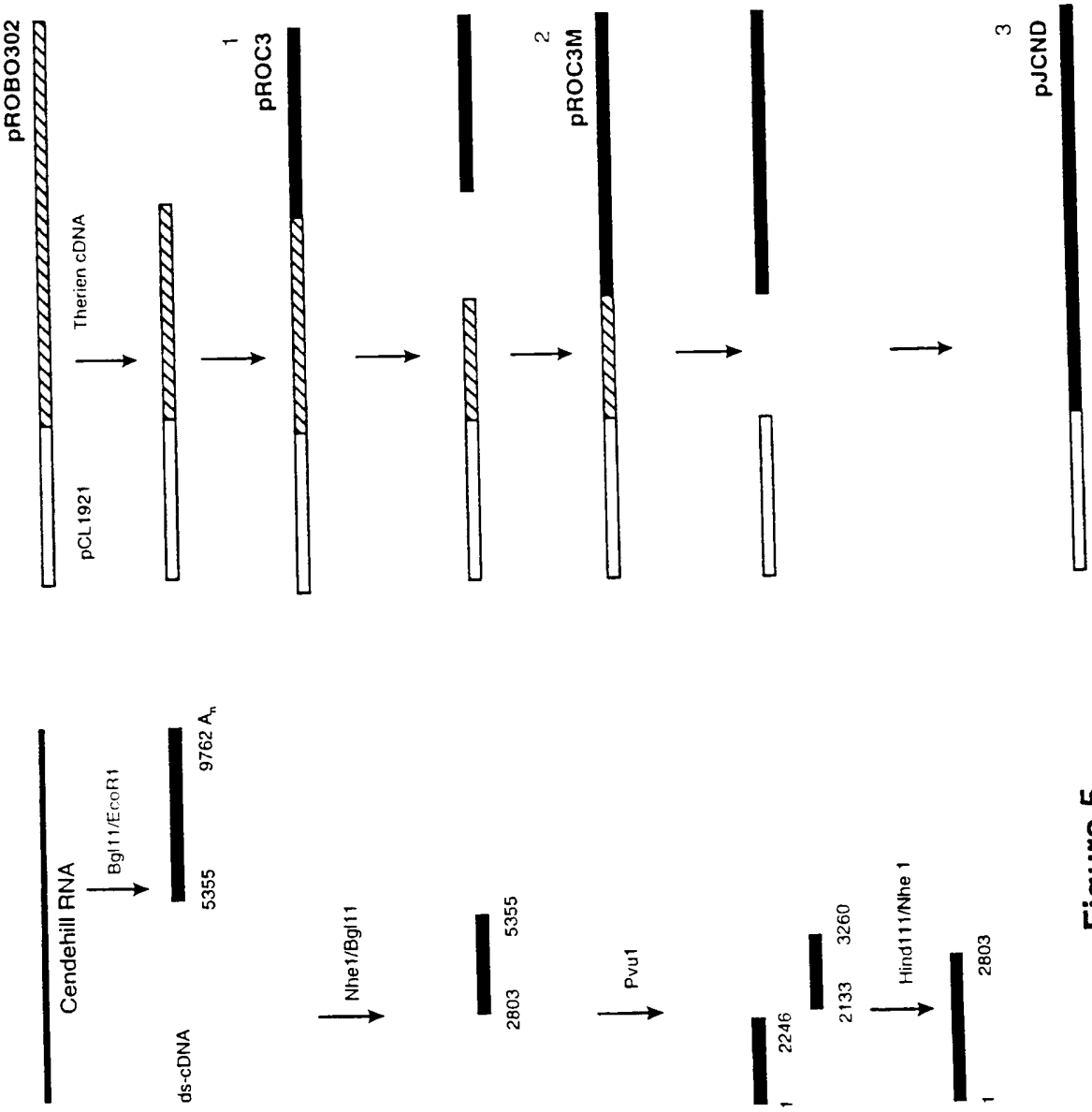


Figure 5

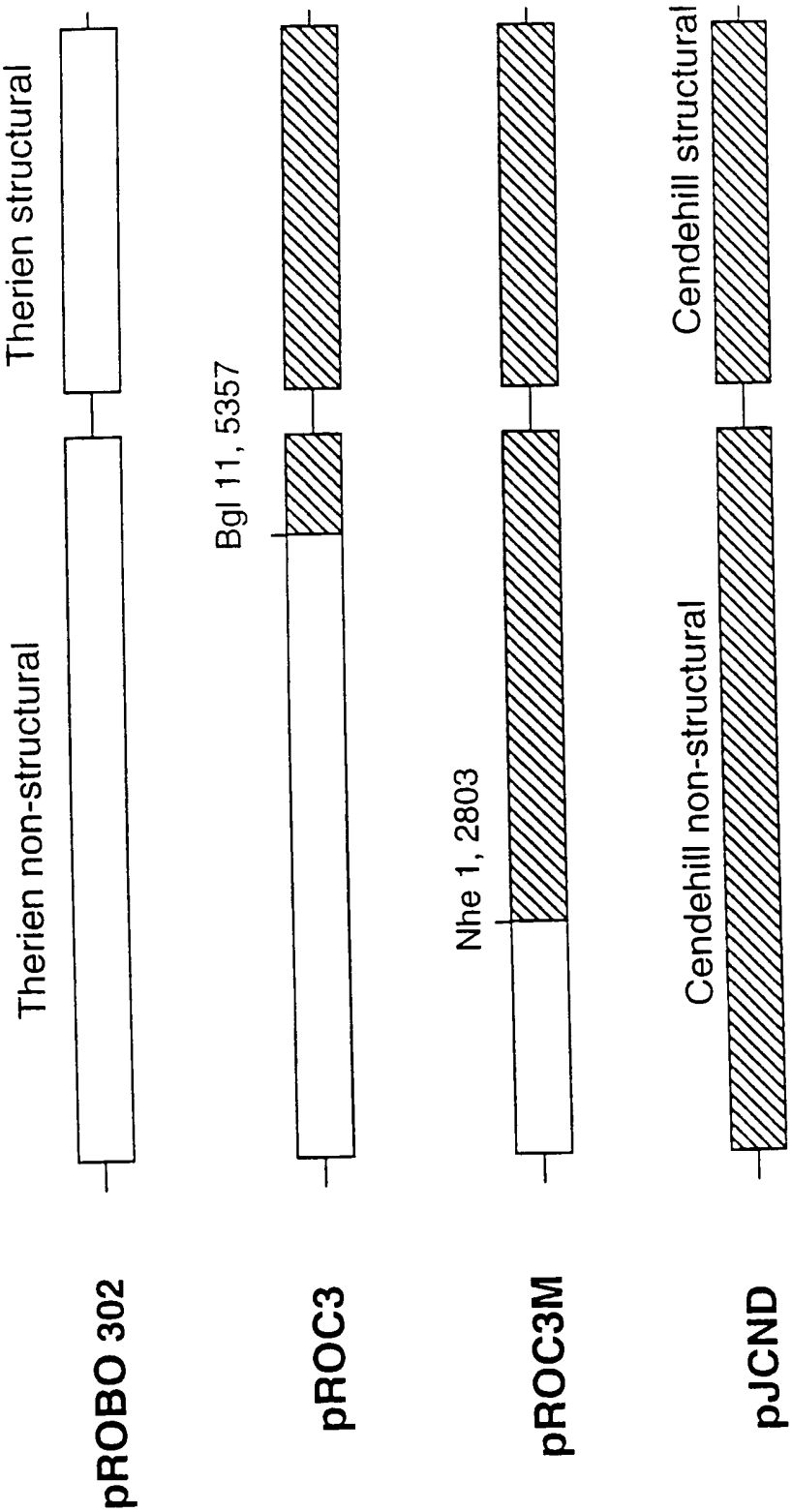


Figure 6

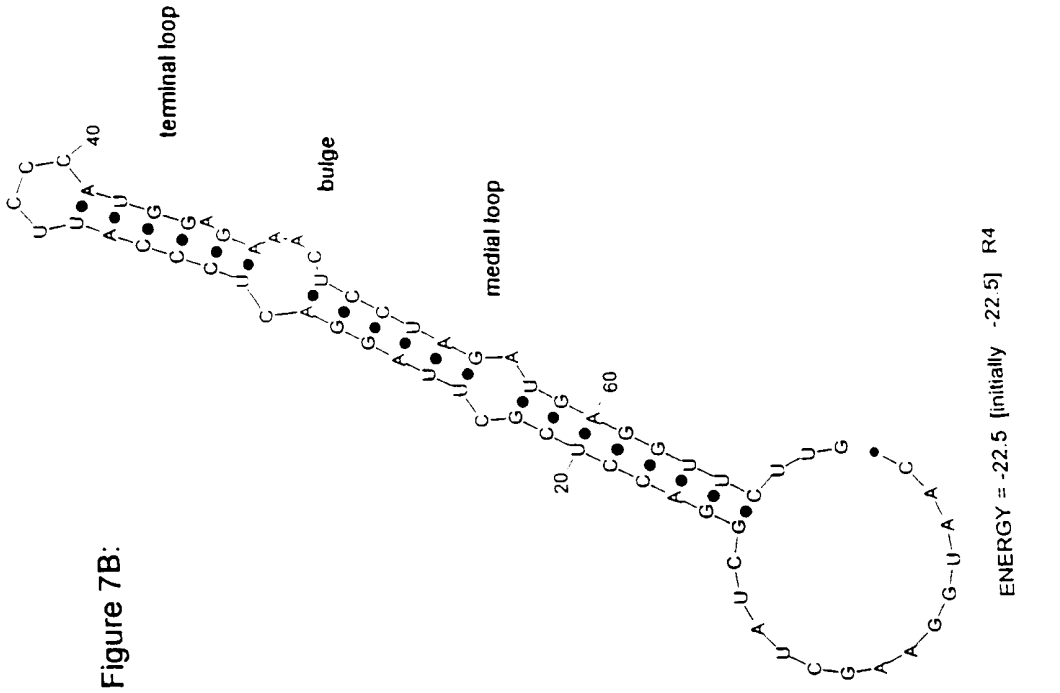


Figure 7B:

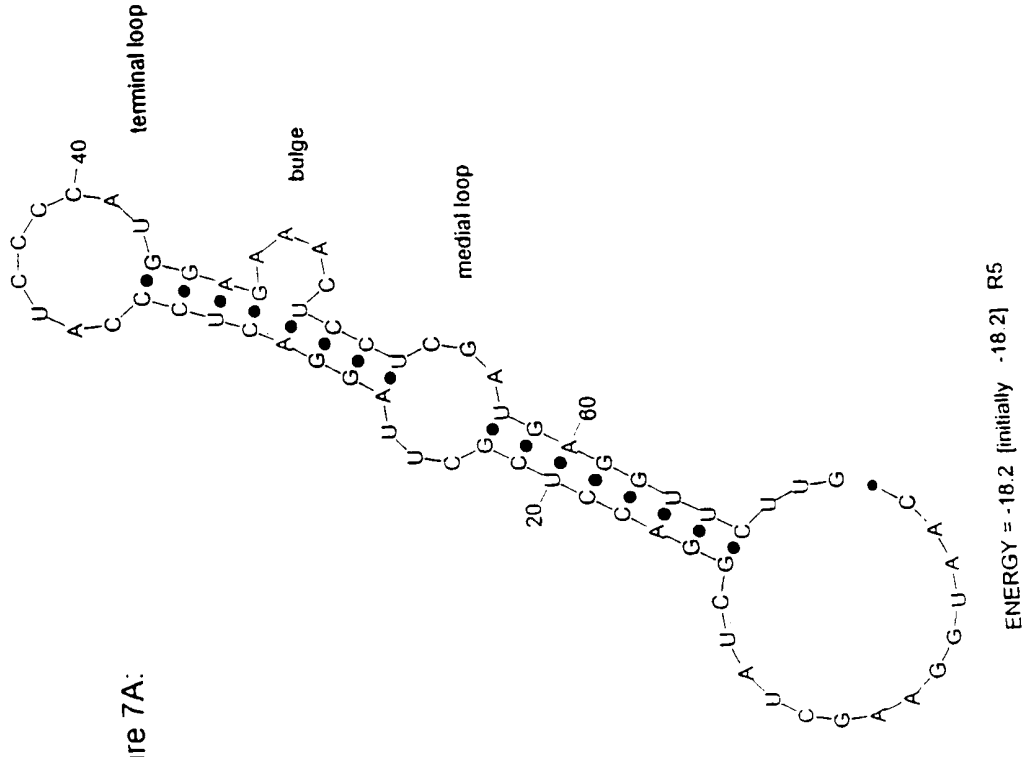
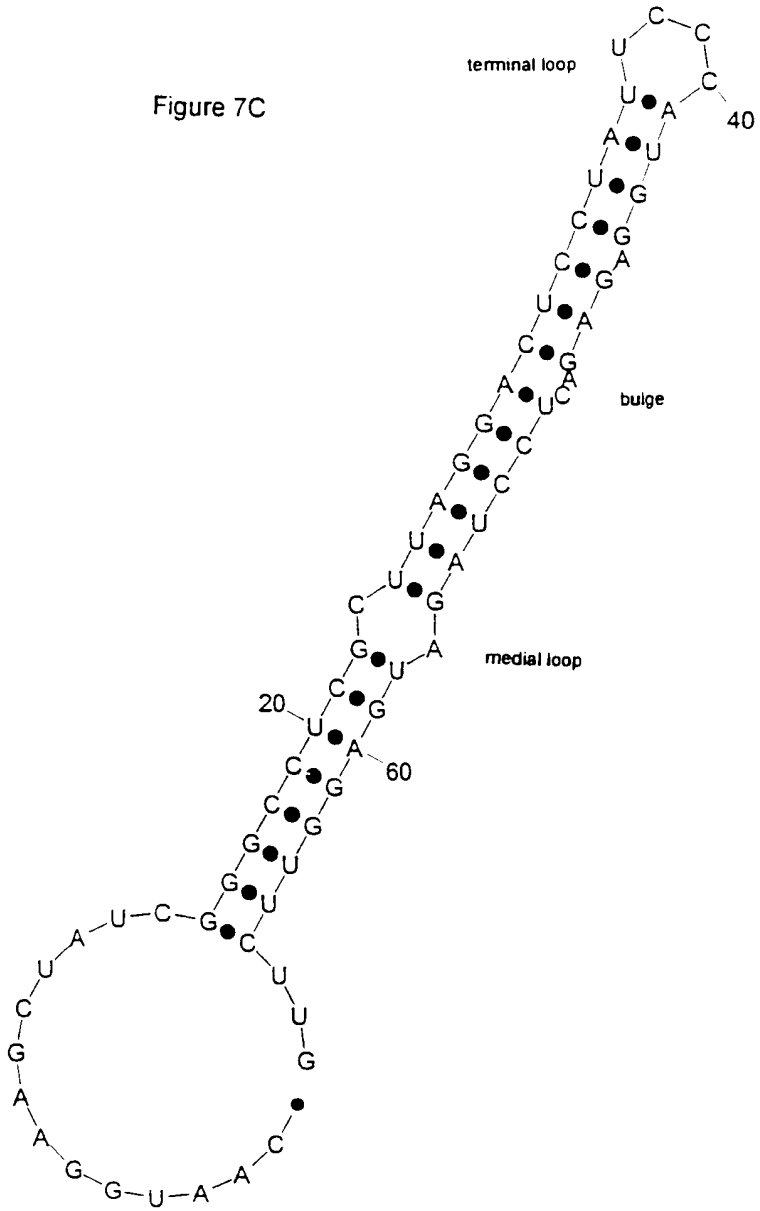


Figure 7A:

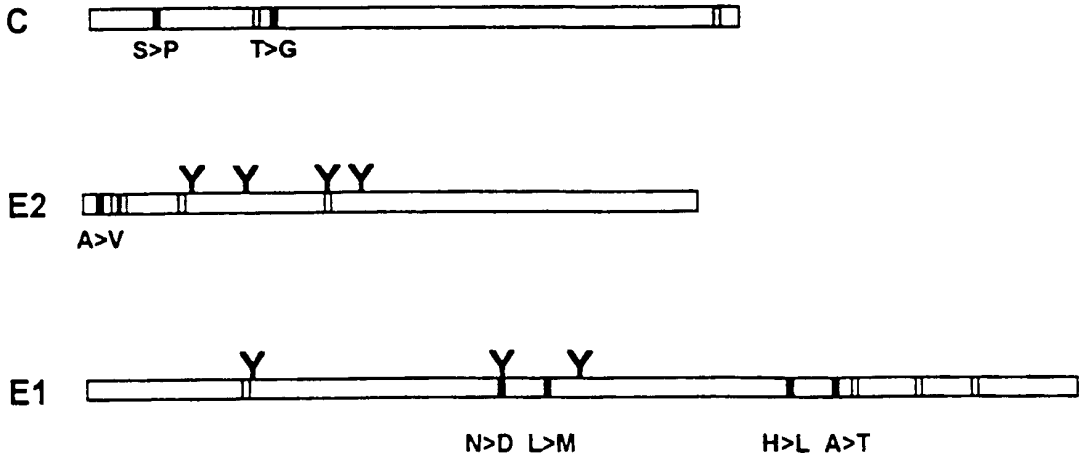
Figure 7C



ENERGY = -22.6 [initially -22.6] R3

▌ Cendehill mutations designated by single letter code

Figure 8



▌ Cendehill

Y Glycosylation site

▯ Cendehill + M33

Figure 9

INTERNATIONAL SEARCH REPORT

Interr onal Application No
PCT/CA 99/00479

A. CLASSIFICATION OF SUBJECT MATTER
IPC 6 C12N15/86 C07K14/19 C12N7/01 C12N7/04 A61K48/00
A61K39/20

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 6 C07K C12N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No
X	BOSMA, T.J. ET AL.: "Nucleotide sequence of a major antigenic domain of the E1 glycoprotein of 22 rubella virus isolates" JOURNAL OF GENERAL VIROLOGY., vol. 77, no. 10, 1996, pages 2523-2530, XP002116125 SOCIETY FOR GENERAL MICROBIOLOGY, READING., GB ISSN: 0022-1317 figure 1; table 1 ---	14
A	US 5 663 065 A (FREY TERYL K ET AL) 2 September 1997 (1997-09-02) the whole document ---	1
A	US 5 439 814 A (FREY TERYL K ET AL) 8 August 1995 (1995-08-08) the whole document ---	1,26
	-/--	

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex

° Special categories of cited documents

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
22 September 1999	05/10/1999

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Chambonnet, F

2

Form PCT/ISA/210 (second sheet) (July 1992)

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	CHANTLER, J. K. ET AL.: "Characterization of rubella virus strain differences associated with attenuation" INTERVIROLOGY, vol. 36, no. 4, 1993, pages 225-236, XP002116126 the whole document	15
A	WO 93 14206 A (CONNAUGHT LAB) 22 July 1993 (1993-07-22) the whole document	1
A	WO 94 03490 A (BIOCHEM IMMUNOSYSTEMS INC) 17 February 1994 (1994-02-17) the whole document	1
A	BE 832 758 A (RECHERCHE ET INDUSTRIE THERAPEUTIQUE , R. I. T.) 26 February 1976 (1976-02-26) the whole document	1
A	CLARKE, D.M. ET AL. : "Nucleotide sequence and in vitro expression of rubella virus 24S subgenomic messenger RNA ancoding the structural proteins E1, E2 and C" NUCLEIC ACIDS RESEARCH, vol. 15, no. 7, 10 April 1987 (1987-04-10), pages 3041-3057, XP002116127 OXFORD GB cited in the application the whole document	1
A	CLARKE, D.M. ET AL.: "Expression of rubella virus cDNA coding for the structural proteins" GENE, vol. 65, 15 May 1988 (1988-05-15), pages 23-30, XP002116128 AMSTERDAM NL the whole document	15
A	NAKHSASI, H. L. ET AL.: "Nucleotide sequence of capsid, E2 and E1 protein genes of Rubella virus vaccine strain RA27/3" NUCLEIC ACIDS RESEARCH., vol. 17, no. 11, 1989, pages 4393-4394, XP002116129 OXFORD UNIVERSITY PRESS, SURREY., GB ISSN: 0305-1048 the whole document	12

-/--

2

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document, with indication where appropriate, of the relevant passages	Relevant to claim No
A	ZHENG, D. ET AL.: "Nucleotide sequence of the 24S subgenomic messenger RNA of a vaccine " GENE: AN INTERNATIONAL JOURNAL ON GENES AND GENOMES., vol. 82, 1989, pages 343-349, XP002116130 ELSEVIER SCIENCE PUBLISHERS, BARKING., GB ISSN: 0378-1119 cited in the application	1,11
P,A	WO 98 20901 A (POUGATCHEV KONSTANTIN V ;FREY TERYL K (US); UNIV GEORGIA (US)) 22 May 1998 (1998-05-22) the whole document	1

INTERNATIONAL SEARCH REPORT

national application No.
PCT/CA 99/00479

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. [X] Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

Remark: Although claims 28 and 29 are directed to a method of treatment of the human/animal body, the search has been carried out and based on the alleged effects of the composition.

2. [ ] Claims Nos.:
because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:

3. [ ] Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. [ ] As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.

2. [ ] As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. [ ] As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:

4. [ ] No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

[ ] The additional search fees were accompanied by the applicant's protest.

[ ] No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

.ormation on patent family members

Intern onal Application No
PCT, CA 99/00479

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 5663065 A	02-09-1997	US 5439814 A	08-08-1995
US 5439814 A	08-08-1995	US 5663065 A	02-09-1997
WO 9314206 A	22-07-1993	AU 662954 B	21-09-1995
		AU 3341393 A	03-08-1993
		CA 2128407 A	22-07-1993
		EP 0624198 A	17-11-1994
		FI 943418 A	15-09-1994
		JP 7503133 T	06-04-1995
		NO 942707 A	20-09-1994
WO 9403490 A	17-02-1994	US 5427792 A	27-06-1995
		AP 453 A	20-01-1996
		AU 690070 B	23-04-1998
		AU 4693493 A	03-03-1994
		CA 2141670 A	17-02-1994
		EP 0656013 A	07-06-1995
		JP 8502724 T	26-03-1996
		ZA 9305642 A	06-02-1995
BE 832758 A	26-02-1976	FR 2321899 A	25-03-1977
		GB 1565658 A	23-04-1980
WO 9820901 A	22-05-1998	AU 5199498 A	03-06-1998